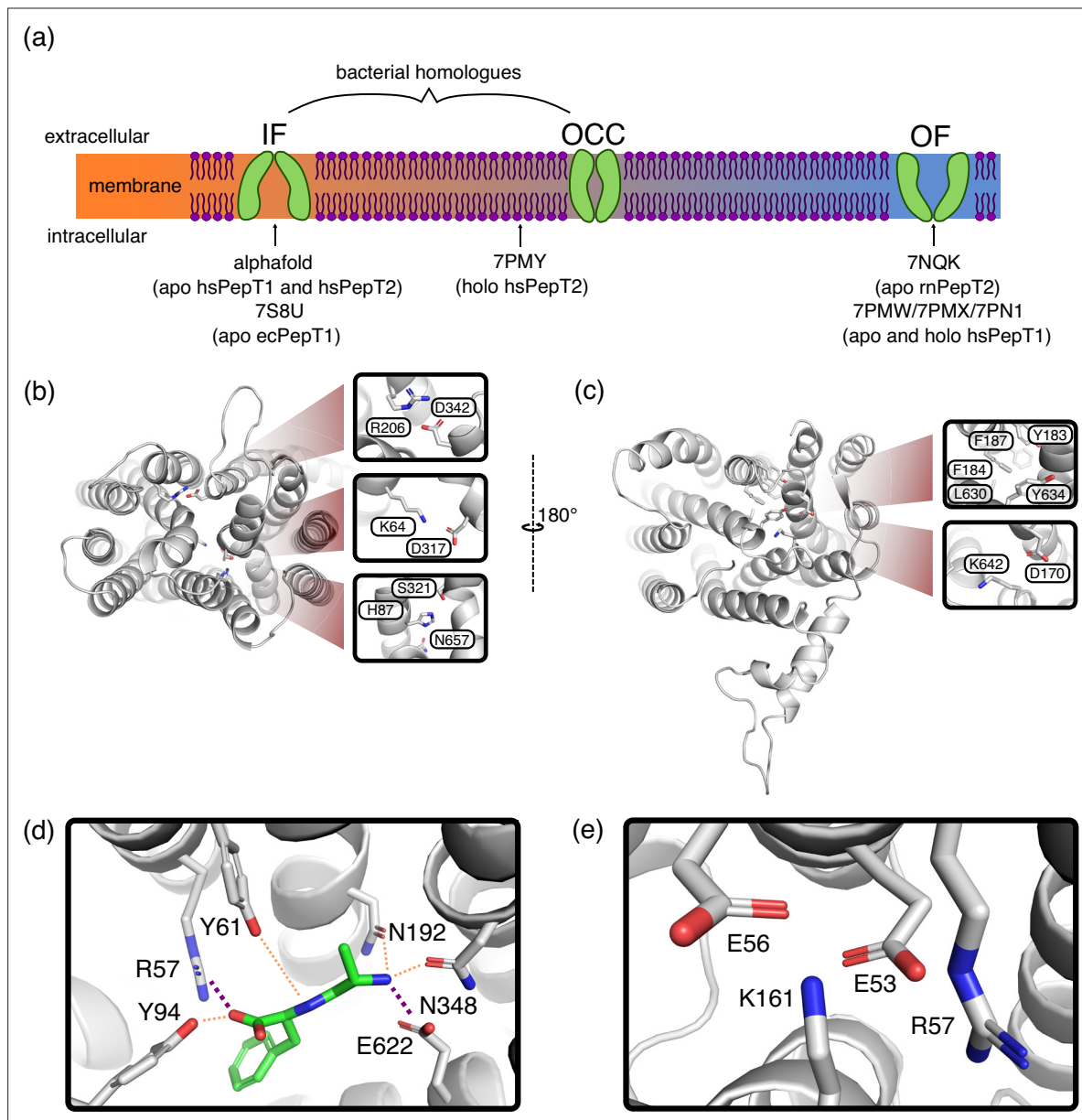


---

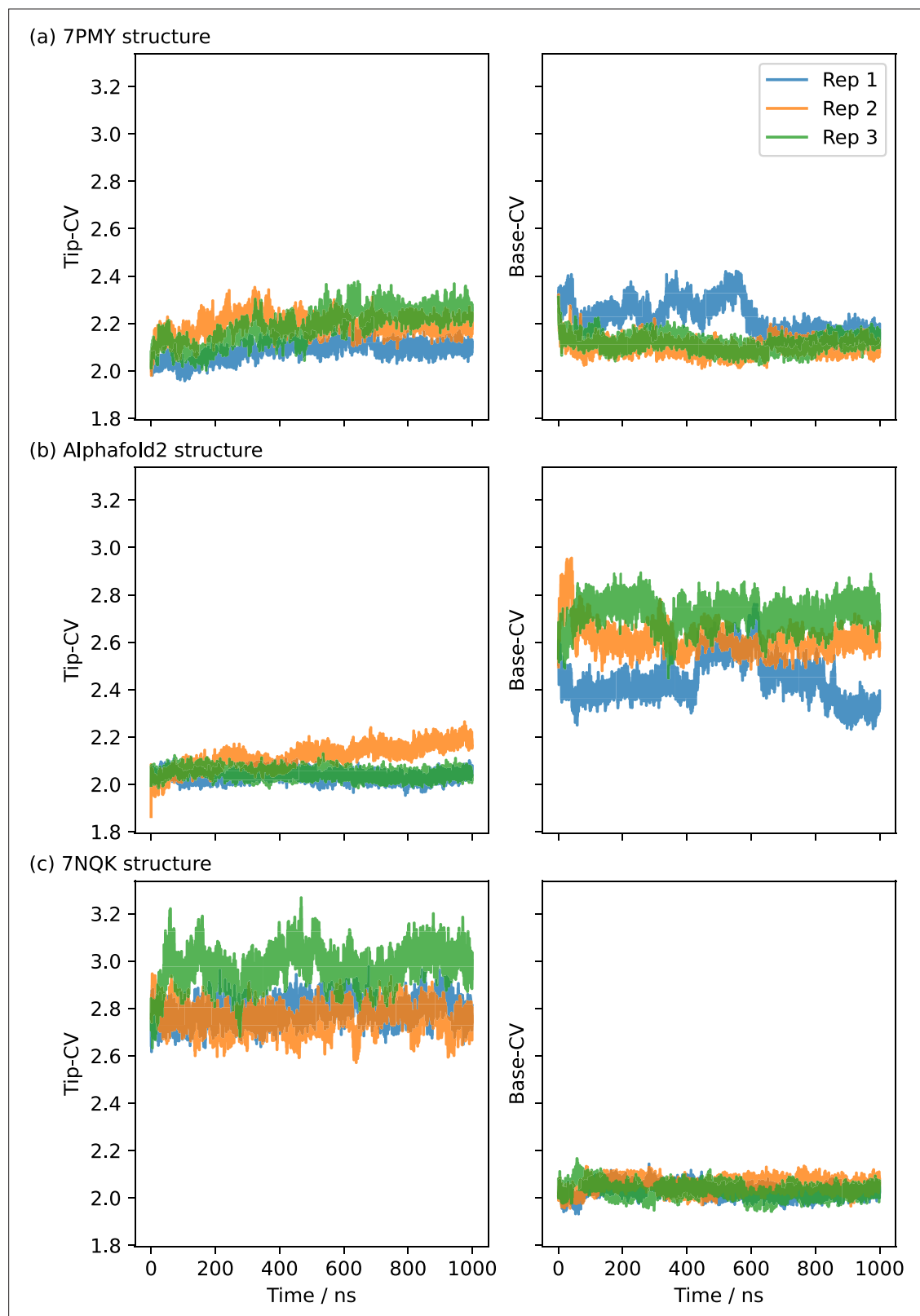
## Figures and figure supplements

The mechanism of mammalian proton-coupled peptide transporters

**Simon M Lichtinger *et al.***



**Figure 1.** Key features of Proton-coupled oligo-peptide transporter (POT) structures. (a) Schematic overview of the conformational diversity of available mammalian POT structures. Intermediate positions between states indicate partial gate opening. (b) AlphaFold-predicted inwards-facing (IF) HsPepT2 structure (top view), highlighting potential inter-bundle extracellular gate interactions. (c) Outwards-facing (OF) Cryo-EM structure of apo RnPepT2 (7NQK, bottom view) (Parker et al., 2021), highlighting potential inter-bundle intracellular gate interactions. (d) Ala-Phe substrate binding pose, representative cluster frame of 1  $\mu$ s molecular dynamics (MD) simulation from 7NQK structure with added ligand, for setup details see Materials and methods. Purple dotted lines represent salt-bridge contacts, orange dotted lines other polar contacts. (e) ExxER motif salt-bridge cluster, representative cluster frame of 1  $\mu$ s MD simulation from 7NQK structure.



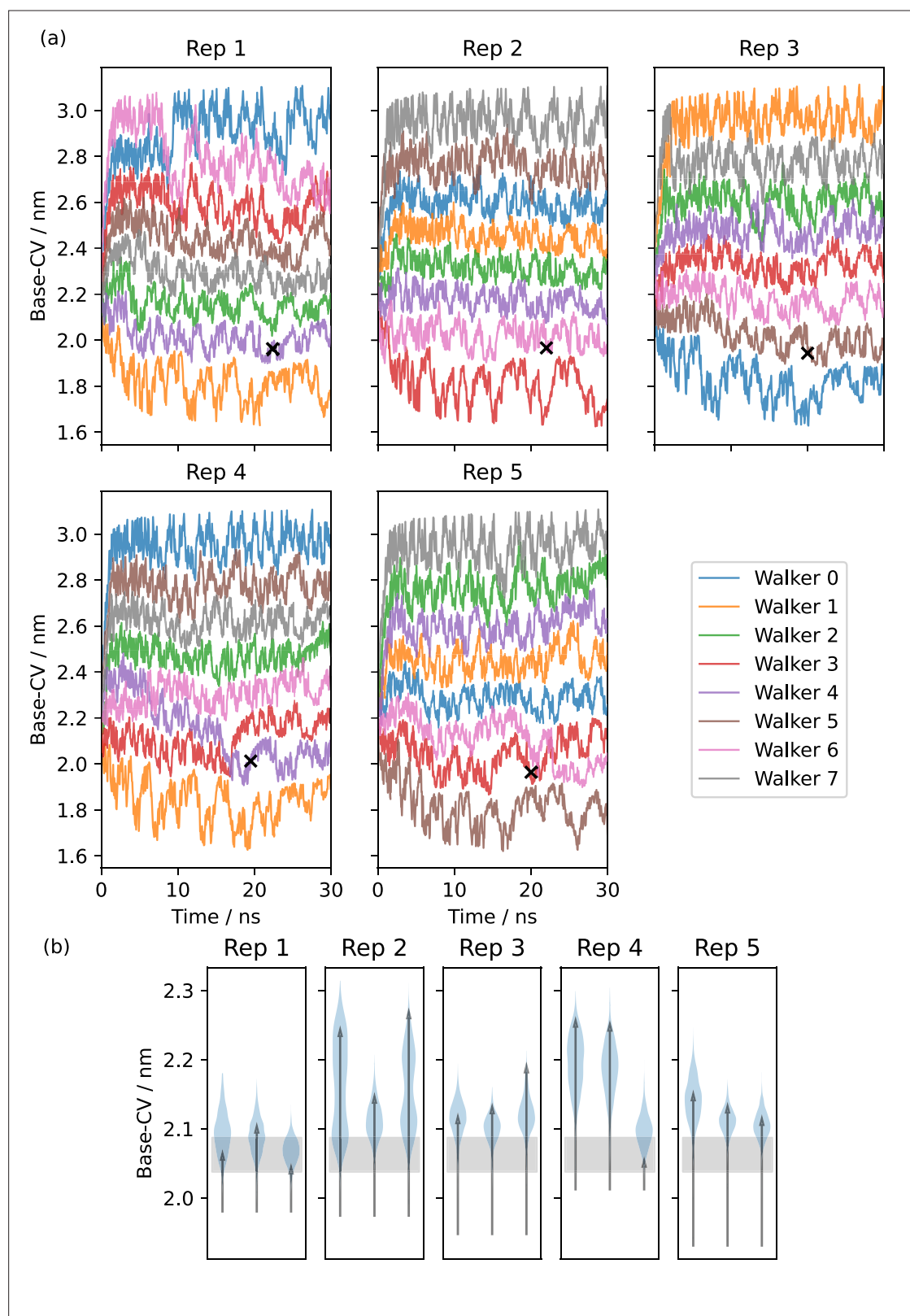
**Figure 1—figure supplement 1.** 1  $\mu$ s-long molecular dynamics (MD) simulations starting from CHARMM-GUI-embedded and equilibrated PepT2 structures. (a) The inwards-facing (IF), partially occluded (OCC) cryo-EM structure (7PMY) moves towards an occluded state via closure of the intracellular gate. However the extracellular gate partially opens in the process. Rep 1 also displays a partial helical unfolding near the intracellular gate (see provided coordinate files in the supplementary data). (b) AlphaFold-based IF embeddings 1 and 3 explore a range of IF conformations while maintaining the structure. (c) AlphaFold-based IF embeddings 1 and 3 explore a range of IF conformations while maintaining the structure.

Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

a stable extracellular gate, whereas rep 2 partially opens the extracellular gate. (c) The outwards-facing (OF) cryo-EM structure (7NQK) remains stable with a tight intracellular gate.



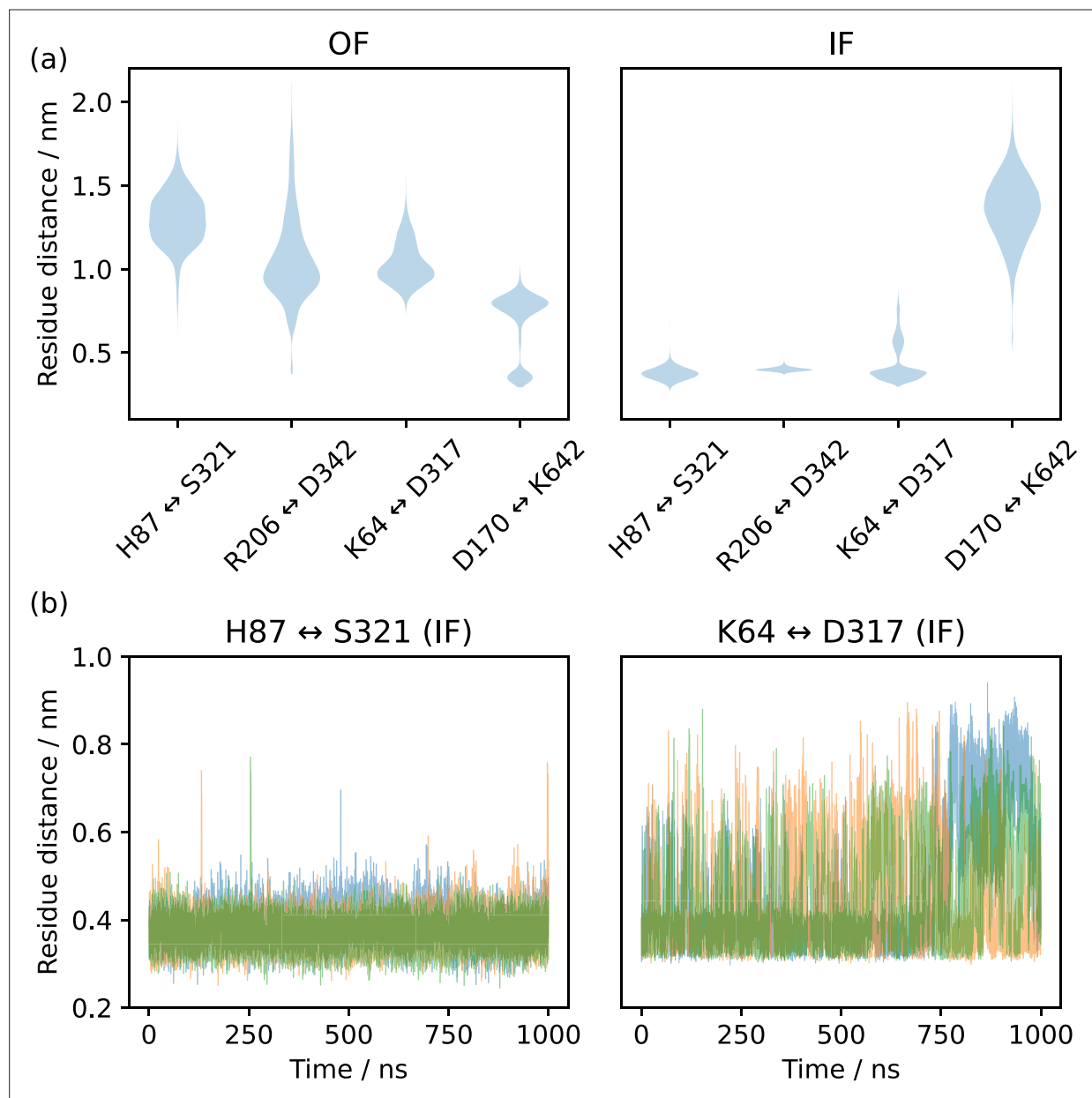


**Figure 1—figure supplement 2.** Metadynamics to derive potential occluded states. **(a)** Five replicates of multiple-walker metadynamics along the Base-CV, starting from the AlphaFold inwards-facing (IF) equilibration rep 1. Black crosses indicate potential occluded (OCC) states picked around a Base-CV value of 2 nm and at a timestep of around 20 ns. **(b)** Equilibrations (three per metadynamics replicate) of the candidate OCC states in 100 ns unbiased molecular dynamics (MD). Histograms of trajectory projections onto the Base-CV are shown as violin plots, arrows link the first and final frames.

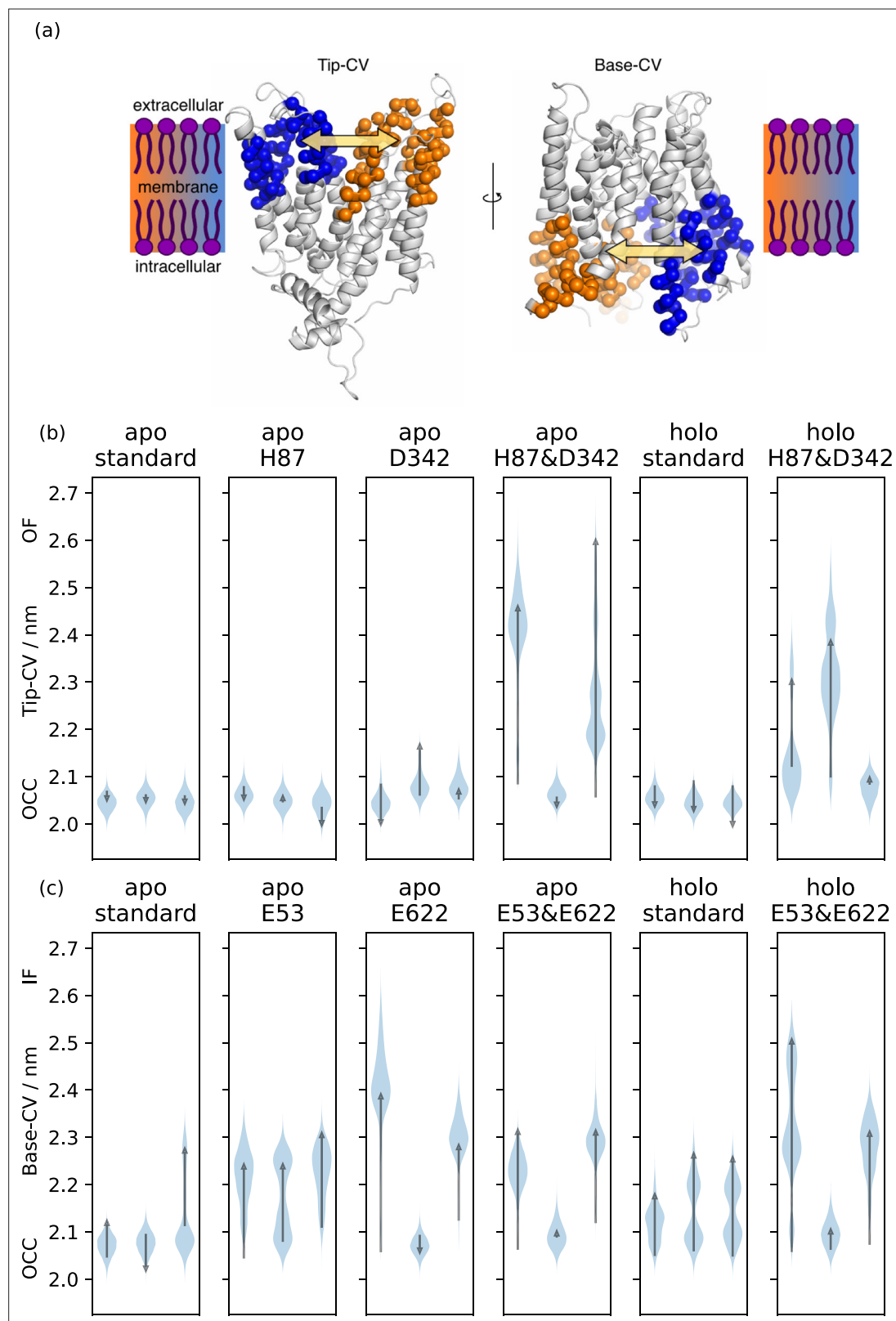
Figure 1—figure supplement 2 continued on next page

*Figure 1—figure supplement 2 continued*

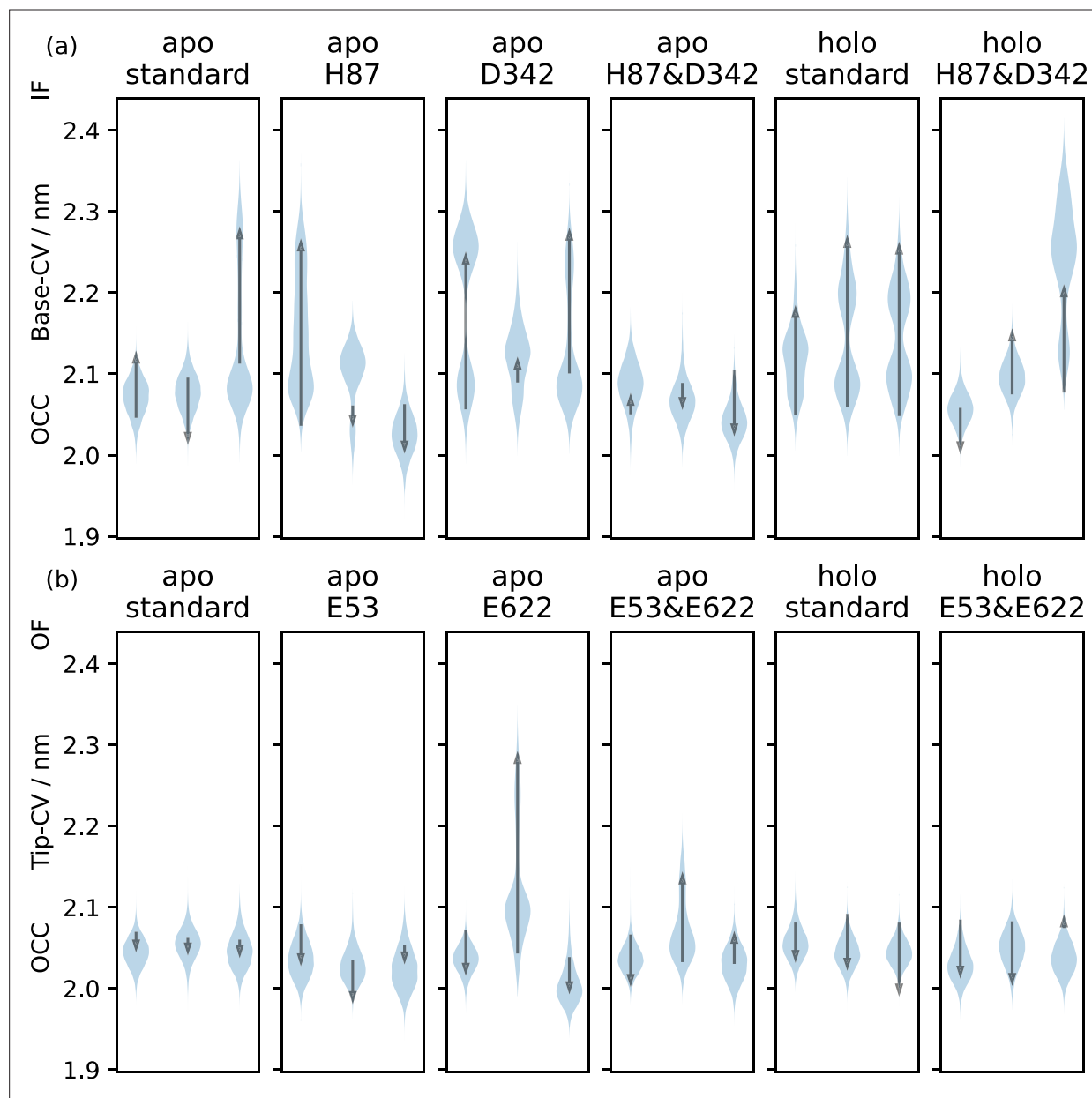
The grey shaded area corresponds to the range of Base-CV values sampled in our outwards-facing (OF) simulation. Rep 1 is a stable OCC state, whereas the other replicates display partial intracellular-gate opening.



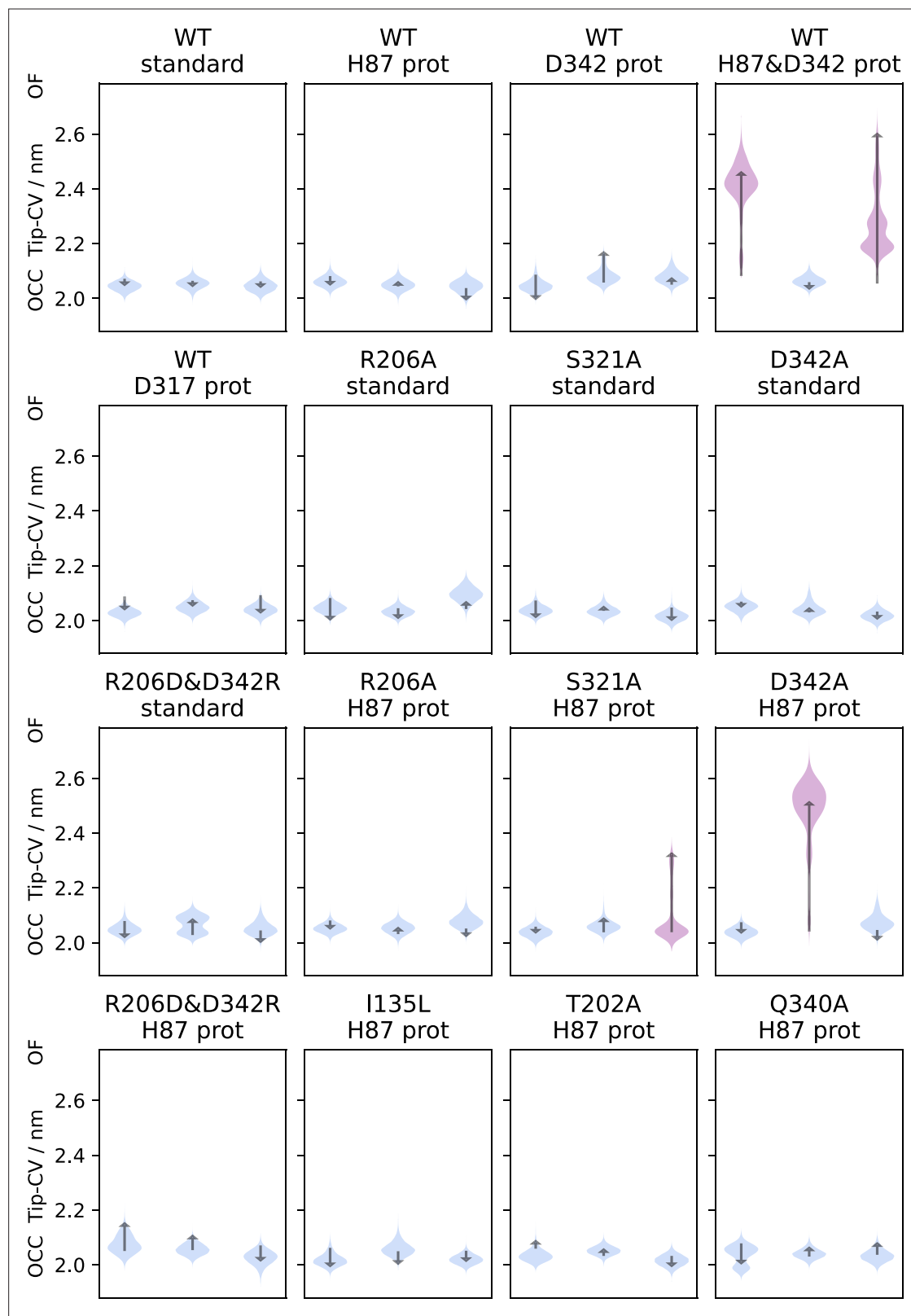
**Figure 1—figure supplement 3.** Inter-residue heavy atom (H87: NE2, S321: OG, R206: CZ, D342: CG, K64: NZ, D317: CG, D170: CG, K642: NZ) distances for several possible gating interactions. (a) Histograms from pooled triplicates of 1  $\mu$ s molecular dynamics (MD) as violin plots in the top row, starting from alphafold-derived inwards-facing (IF) and Cryo-EM derived outwards-facing (OF) conformations. H87↔S321 and R206↔D342 are always formed in the IF state but not in the OF state, while K64↔D317 and D170↔K642 show a preference for the IF and OF states, respectively but are not formed in all trajectory frames. (b) Triplicate time series of the H87↔S321 and K64↔D317 interactions in the IF state, showing how the former is a tight interaction, while the latter is unstable and only transiently formed.



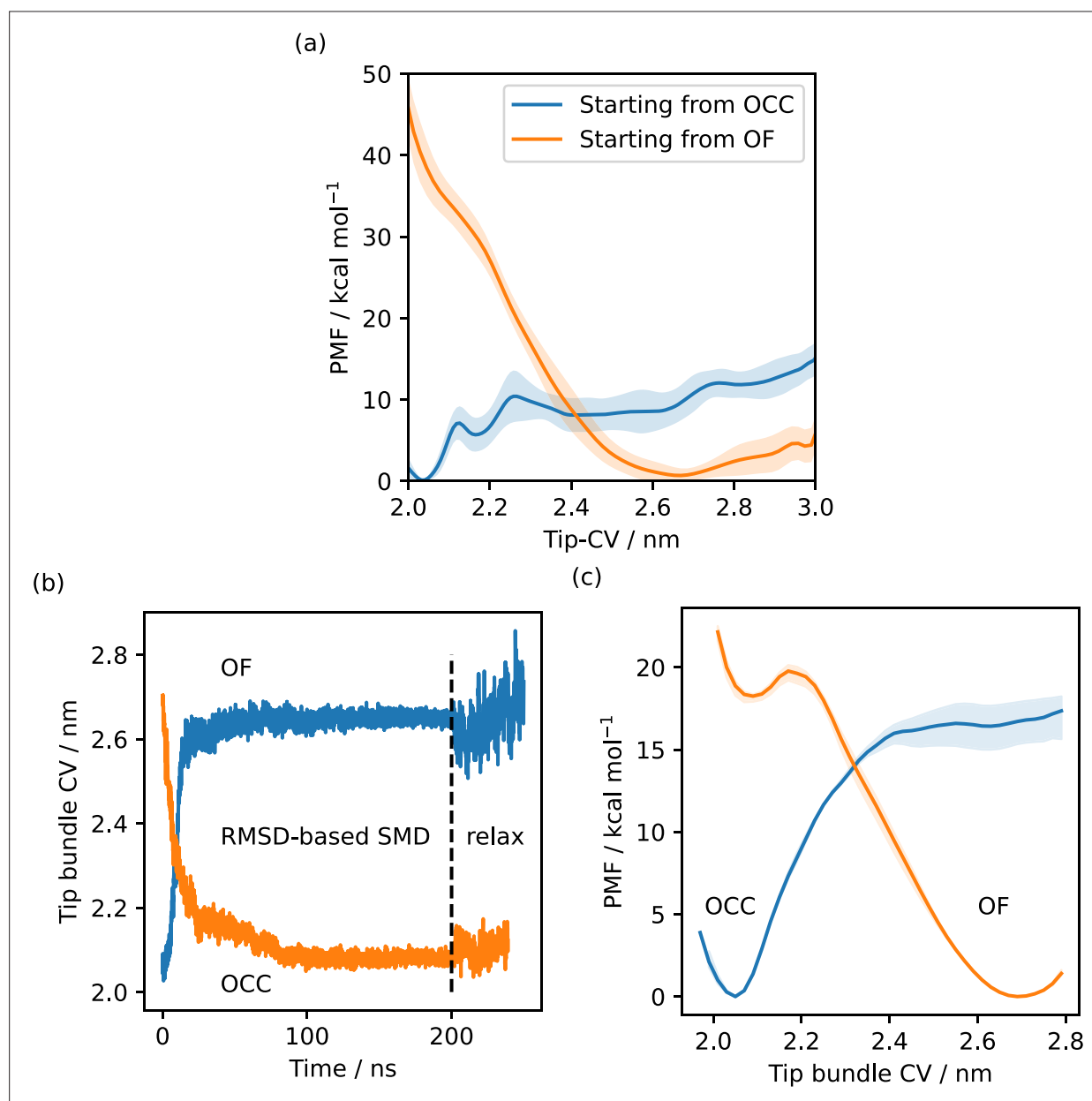
**Figure 2.** Collective variable analysis. **(a)** Illustration of the collective variables (CVs) used to quantify extra- and intracellular gate opening, consisting of inter-bundle centre-of-mass distances between the helical tips (top 11 residues) and bases (bottom 11 residues). **(b)** + **(c)** Triplicate 1  $\mu$ s-molecular dynamics (MD) simulations starting from occluded (OCC), showing the effects of different protonation and substrate binding states, projected onto the **(b)** Tip-CV and **(c)** Base-CV, respectively. Violin plots are trajectory histograms, arrows link the CV values of the first and last frames.



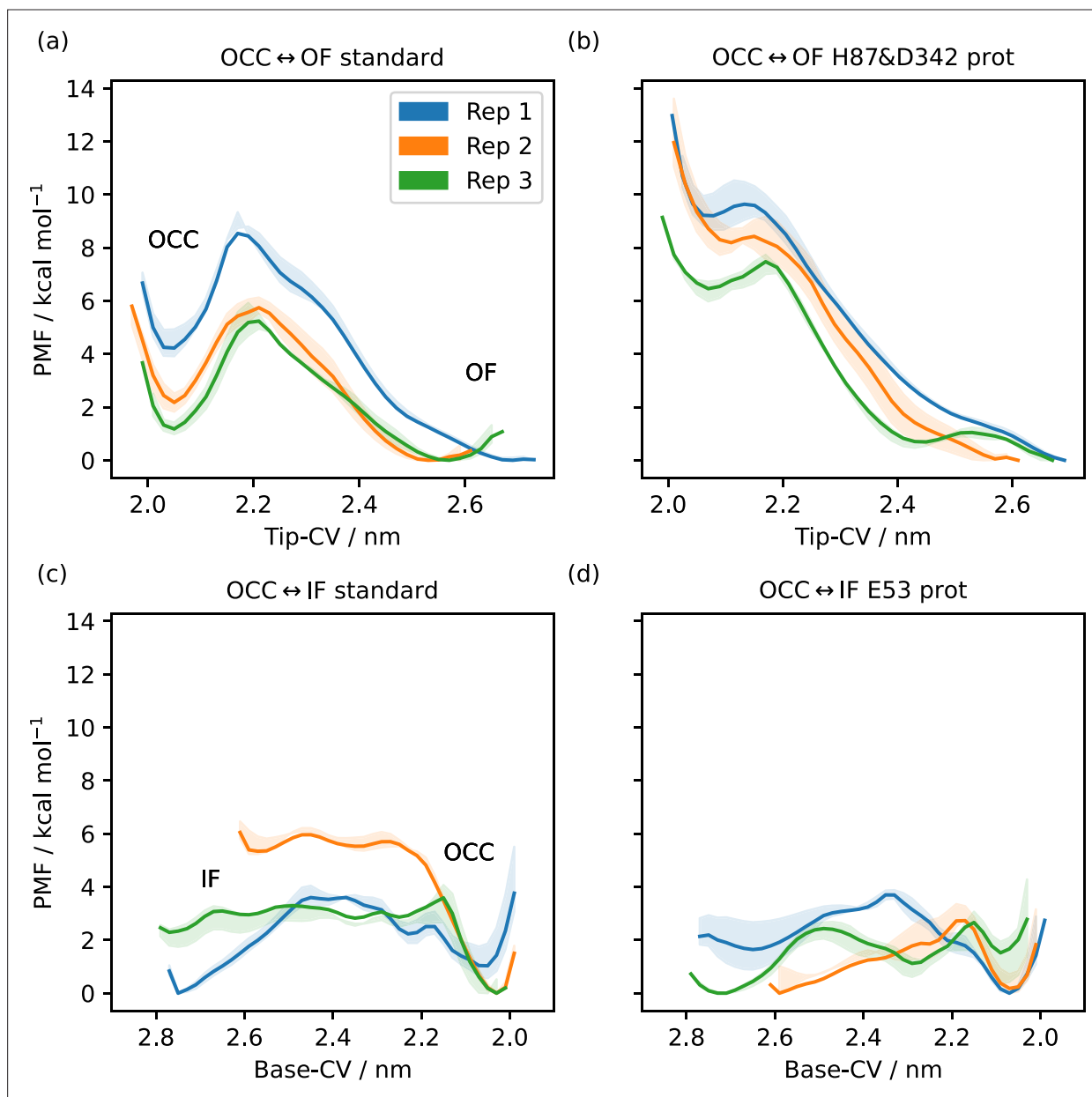
**Figure 2—figure supplement 1.** Effects of protonation and substrate on vanilla MD simulations of the OCC state. Triplicate 1  $\mu$ s-molecular dynamics (MD) simulations starting from occluded (OCC), showing the effects of different protonation and substrate binding states, projected onto the (a) Base-collective variables (CV) and (c) Tip-CV, respectively. Violin plots are trajectory histograms, arrows link the CV values of the first and last frames. Intracellular gate flexibility is suppressed by conditions that favour extracellular gate opening and vice versa.



**Figure 2—figure supplement 2.** Triplicate 1  $\mu$ s-molecular dynamics (MD) simulations starting from the occluded (OCC) state, showing the effects of different protonation states and mutations projected onto the tip-collective variables (CV). Violin plots are trajectory histograms, arrows link the CV values of the first and last frames. Trajectories which displayed significant extracellular gate opening are highlighted in purple. Spontaneous extracellular gate opening requires H87 protonation, and the disruption of the R206↔D342 salt bridge also makes a significant contribution, either by mutation or protonation.

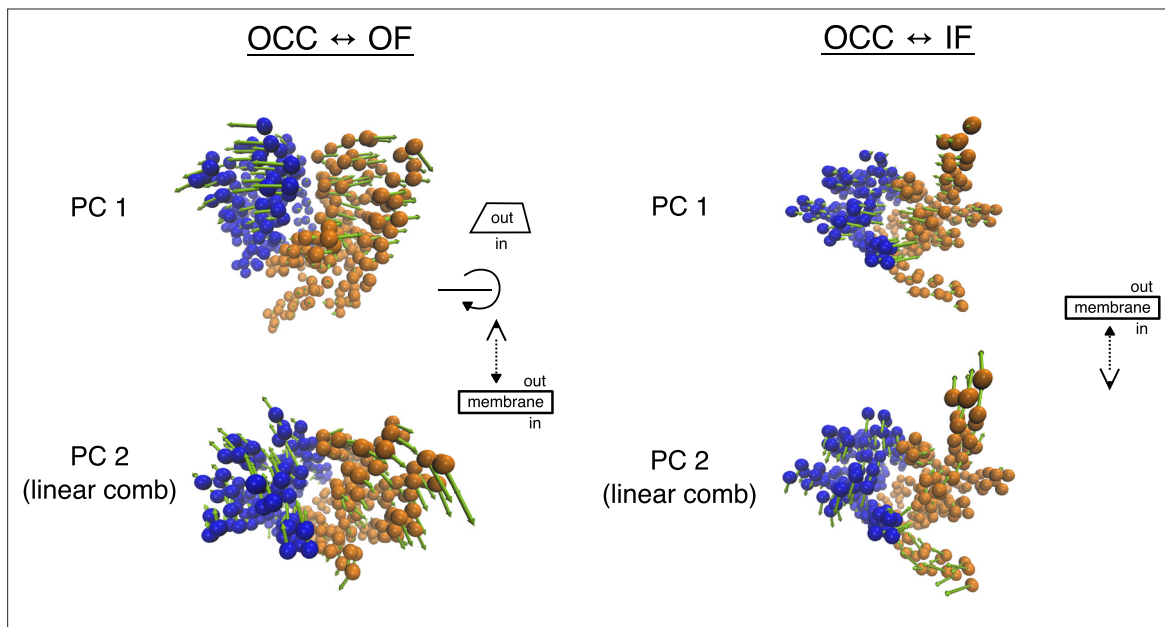


**Figure 2—figure supplement 3.** Illustration of hysteresis effects. (a) Free energy profiles from metadynamics simulations (eight walkers) along the Tip-collective variables (CV), starting simulations from the occluded (OCC) (total sampling 1.7  $\mu$ s) or outwards-facing (OF) state (total sampling 860 ns). Solid lines are the free energy estimates using the second half of the data only, shaded area is the standard deviation of free energy estimates with respect to sequential data chunks. The disparity between the curves indicates a significant hysteresis problem, favouring the initial state of the respective metadynamics run. (b) Steered MD (SMD) runs between OF and OCC states, biasing the heavy-atom replica-exchange umbrella sampling (RMSD) to the respective target state, shown as projections along the tip-CV. Metastable OCC and OF states are formed that remain stable in 50 ns unbiased molecular dynamics (MD). (c) REUS along the tip-CV with starting conformations picked to be equidistant in the CV from two SMD runs. Sampling was using 48 windows for a total of 4.4  $\mu$ s (OCC→OF path) and 6.1  $\mu$ s (OF→OCC path). Solid lines are PMFs calculated using all sampling, the shaded areas are error ranges obtained by omitting either the first 40% or the last 40% of sampling. The disparity between the curves indicates a significant hysteresis problem, favouring the initial state of the respective SMD path-generation run.

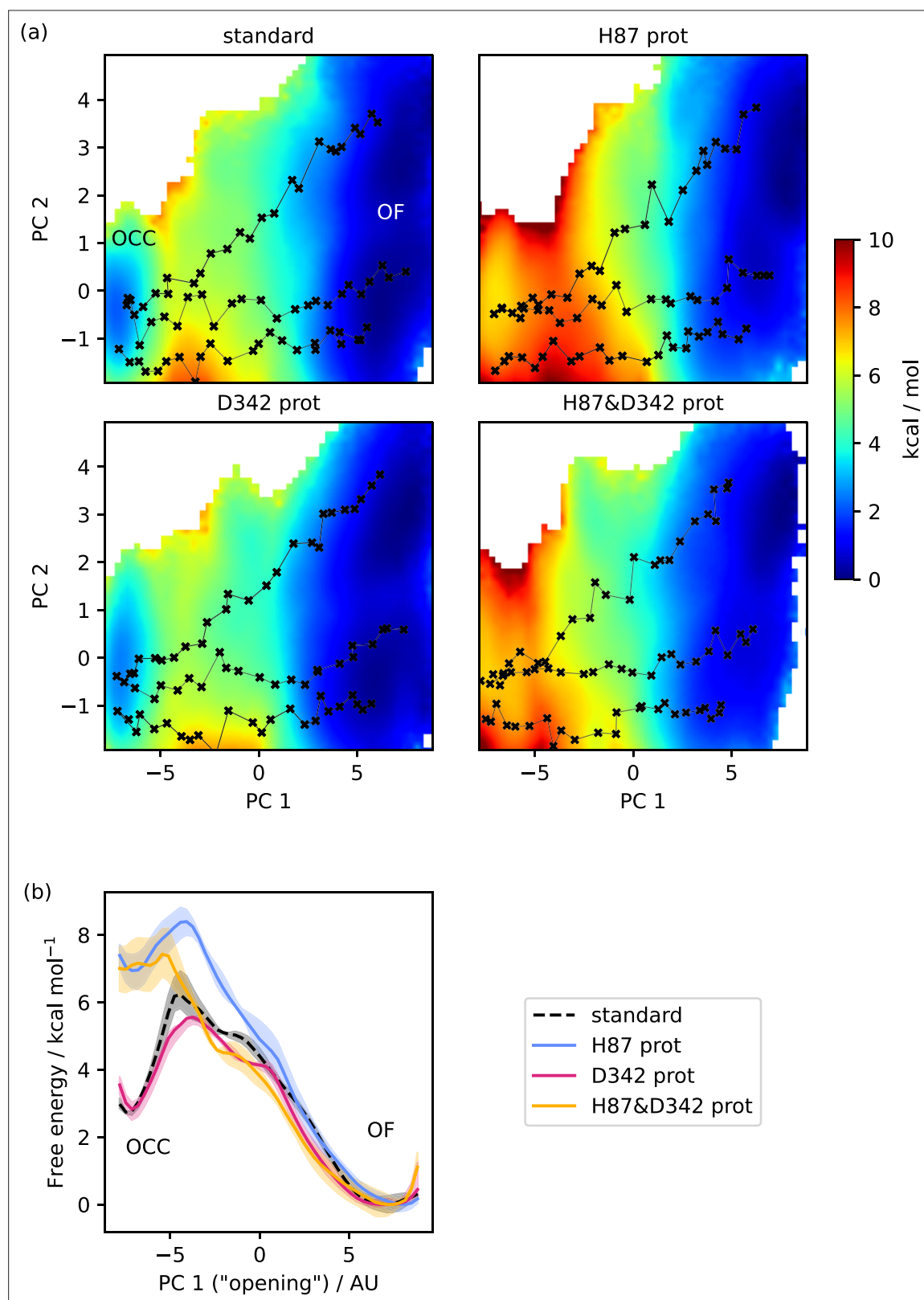


**Figure 2—figure supplement 4.** 1D-potential of mean forces (PMFs) along the Tip-collective variable (CV) or Base-CV (as indicated), from replica-exchange umbrella sampling (REUS) starting at Morphing Endstates by Modelling Ensembles with iNdependent TOpologies (MEMENTO) intermediates. The central line is calculated using all sampling, whereas the shaded areas are enclosed by curves derived from omitting the first 40% or the last 40% of sampling, giving a sense of apparent convergence (while substantial inter-replicate differences remain). (a) OCC ↔ OF transition, standard protonation states. Distinct occluded (OCC) and outwards-facing (OF) basins with separating barrier. (b) OCC ↔ OF transition, H87 and D342 protonated. The OCC basin and the separating barrier largely disappear. (c) OCC ↔ IF transition, standard protonation states. Distinct OCC basin and raised inwards-facing (IF) plateau. (d) OCC ↔ IF transition, E53 protontaed. The IF plateau and the barrier are lowered with respect to OCC.





**Figure 2—figure supplement 5.** Illustration of the principal component analysis (PCA)-derived collective variables (CVs) for 2D-replica-exchange umbrella sampling (REUS). TM-helix CA-atoms are shown as spheres, coloured blue for the N-terminal bundle and orange for the C-terminal bundle, while arrows show magnitude and direction. For the OCC↔OF transition, PC 1 corresponds to the gating motion, while PC 2 is a cleft-sliding movement. For OCC↔IF, PC 1 corresponds to the gating motion, while PC 2 is a twisting movement. See **Videos 1–4** for animated versions of the same representations.

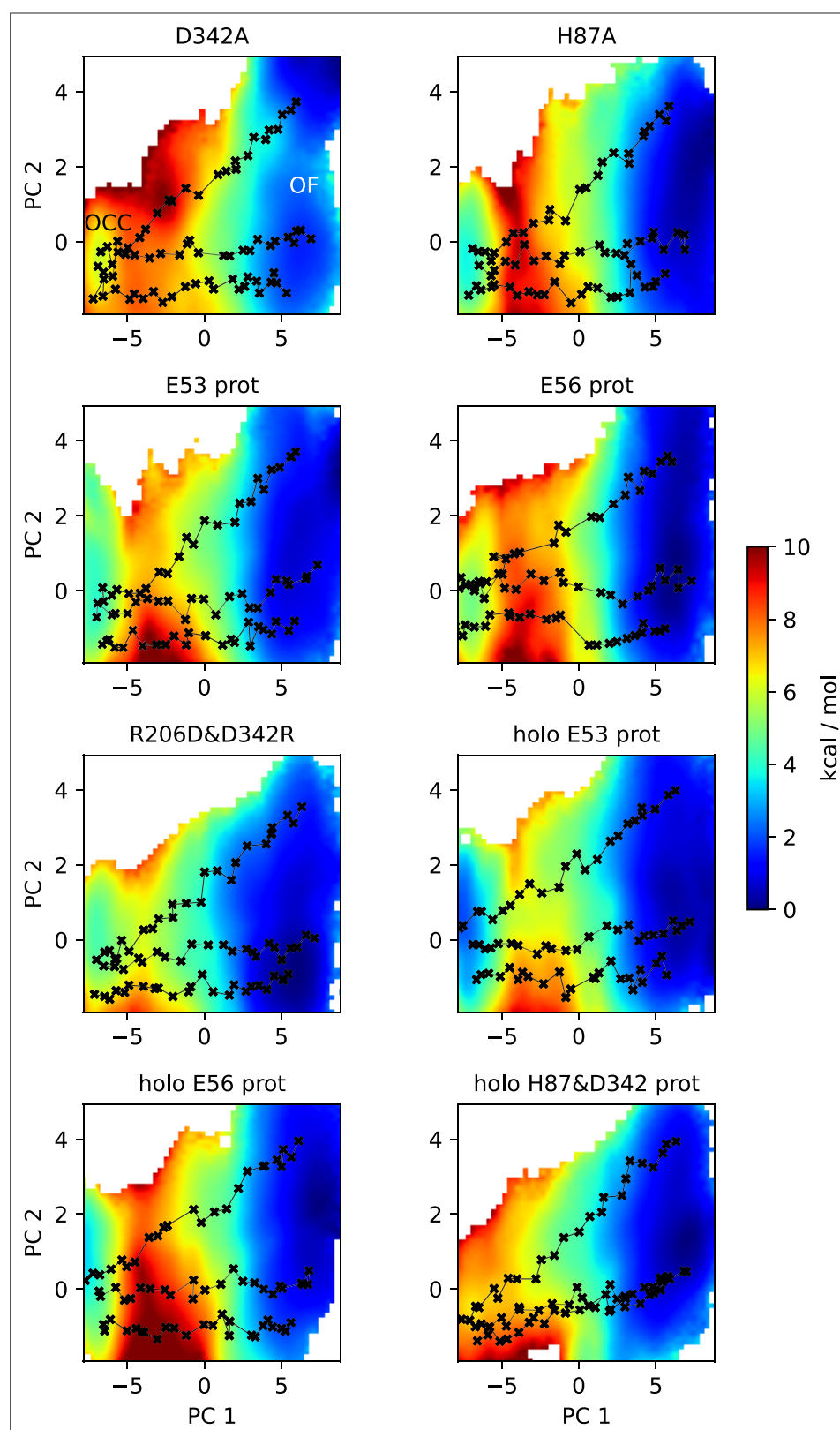


**Figure 3.** Free energy surfaces for the OCC-OF transition. (a) 2D-PMFs from replica-exchange umbrella sampling (REUS) starting with Morphing Endstates by Modelling Ensembles with iNdependent TOpologies (MEMENTO) paths, in different protonation states of candidate extracellular gating residues. (b) Projection of the 2D-PMFs in part a onto PC 1 using Boltzmann reweighting. Shaded areas indicate convergence errors as the range of PMF values for a given CV value obtained with the first 40%, the last 40%, and 100% of sampling included (after alignment to the 100% curve). H87 and

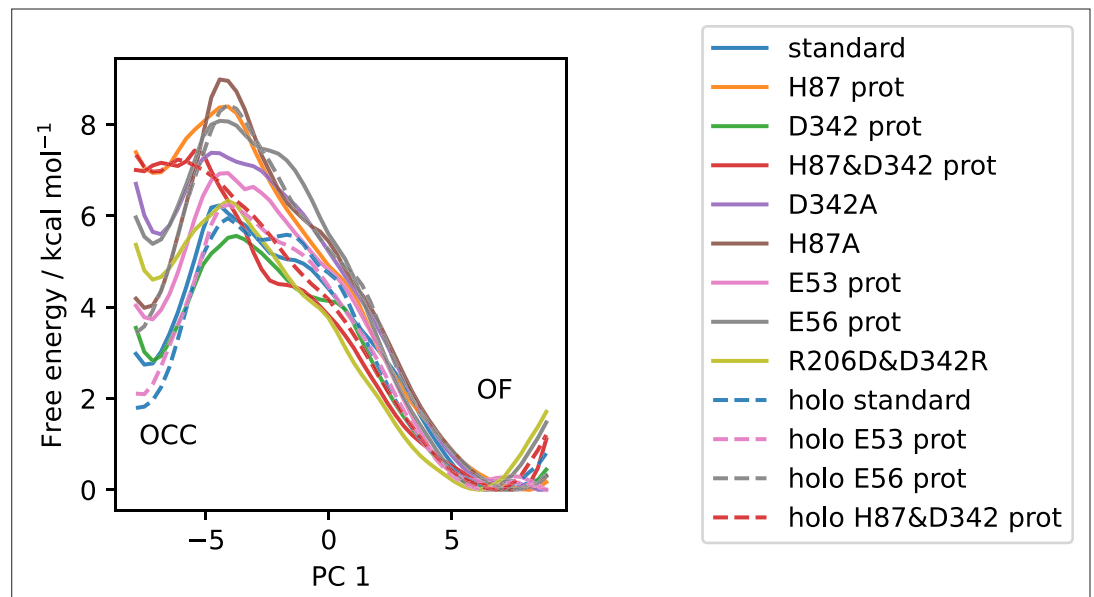
*Figure 3 continued on next page*

*Figure 3 continued*

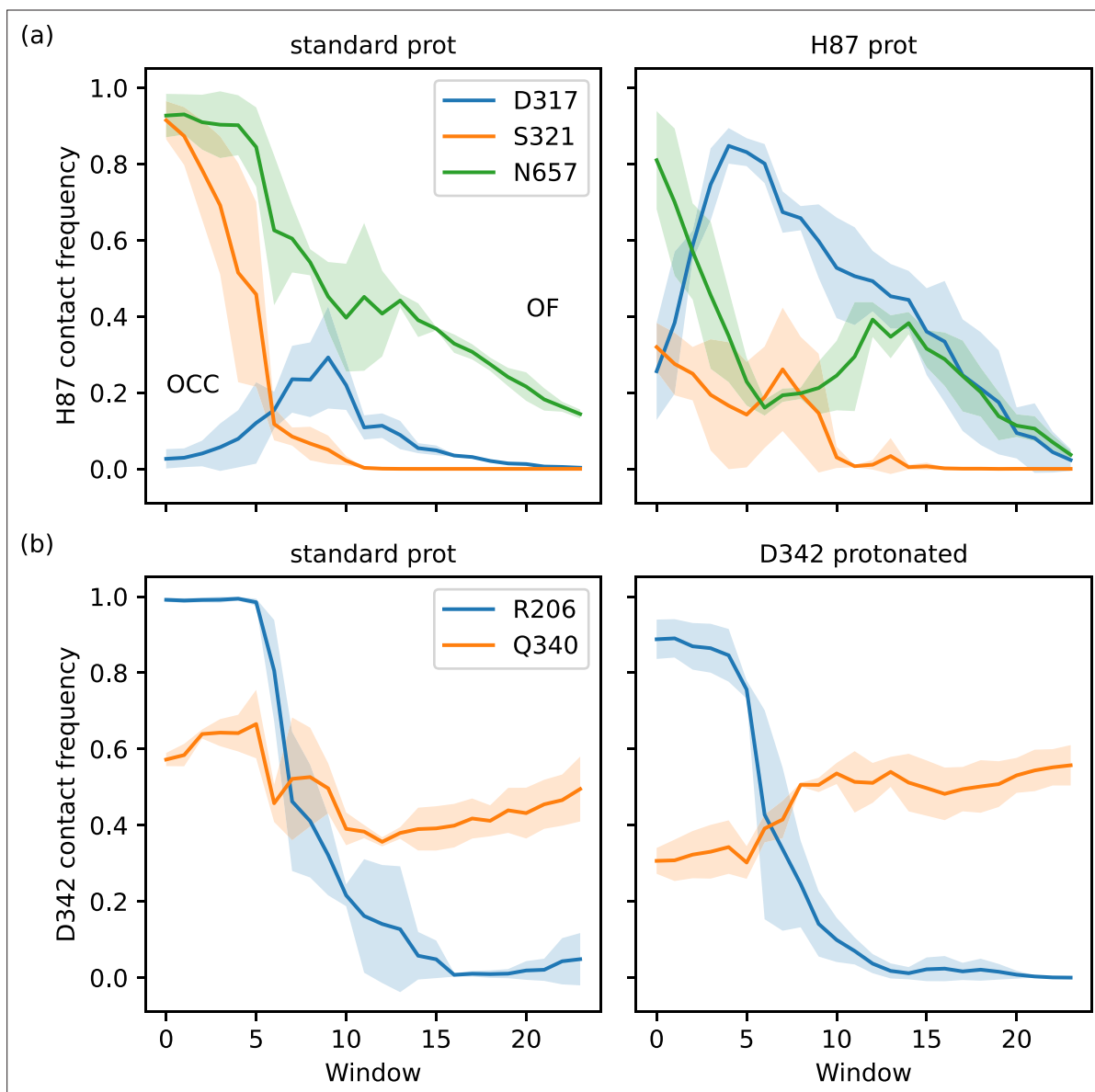
D342 form an additive extracellular gate, where H87 protonation changes the relative occluded (OCC)–outwards-facing (OF) state energies as well as the transition barrier, while D342 protonation only contributes in the transition region. Note that the individual PMFs are only determined by our REUS approach up to additive constants, and are shown aligned here at the OF state for convenience of comparison.



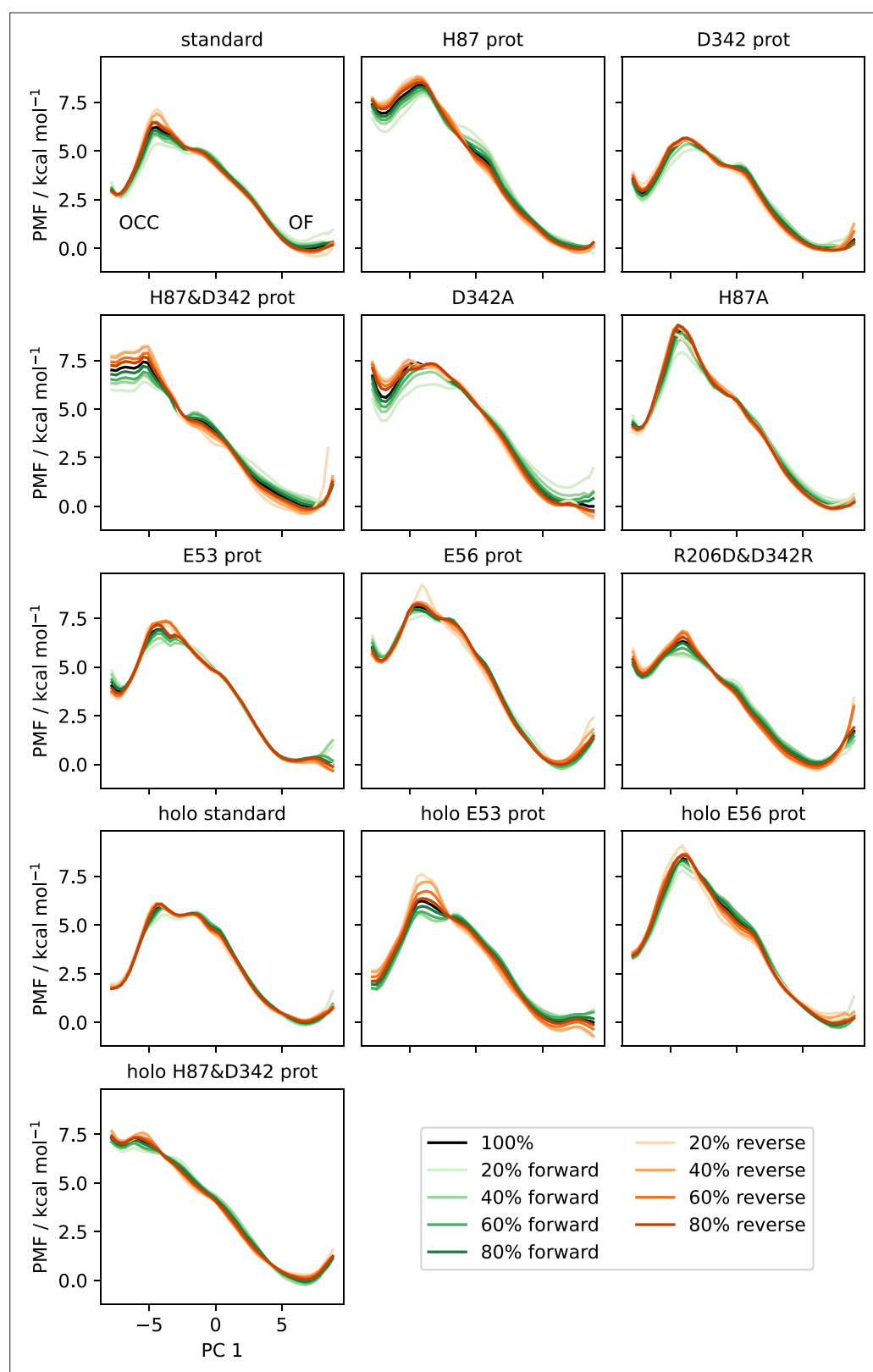
**Figure 3—figure supplement 1.** 2D-PMFs of the OCC→OF transition from replica-exchange umbrella sampling (REUS) with Morphing Endstates by Modelling Ensembles with iNdependent TOPologies (MEMENTO) paths in additional protonation state and mutation conditions.



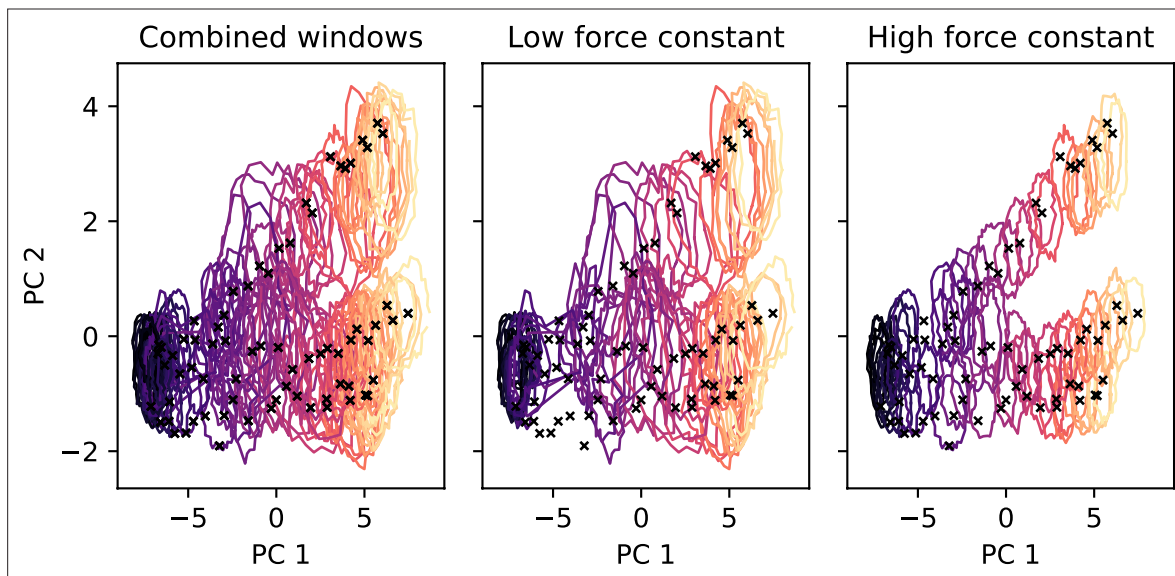
**Figure 3—figure supplement 2.** All available OCC→OF 2D-PMFs, projected onto the first collective variable (CV) (PC 1). Solid lines are apo PMFs, dashed lines are ala-phe substrate-bound and color-matched to the respective apo PMF. Note that the individual PMFs are only determined by our replica-exchange umbrella sampling (REUS) approach up to additive constants, and are shown aligned here at the outwards-facing (OF) state for convenience of comparison.



**Figure 3—figure supplement 3.** Interaction plots of the 2D-PMF trajectory data (high force constant windows only), calculated as frequencies of finding inter-residue heavy-atom distances smaller than 0.35 nm, shown as a line for the average across three replicates with shaded standard deviations. (a) H87 interactions with D317, S321, and N657. Protonation of H87 replaces the S321 interaction by an interaction with D317. (b) D342 interactions with R206 and Q340. The tight salt bridge D342–R206 is disrupted by D342 protonation, but the residues still interact in the occluded (OCC) state via hydrogen bonds.

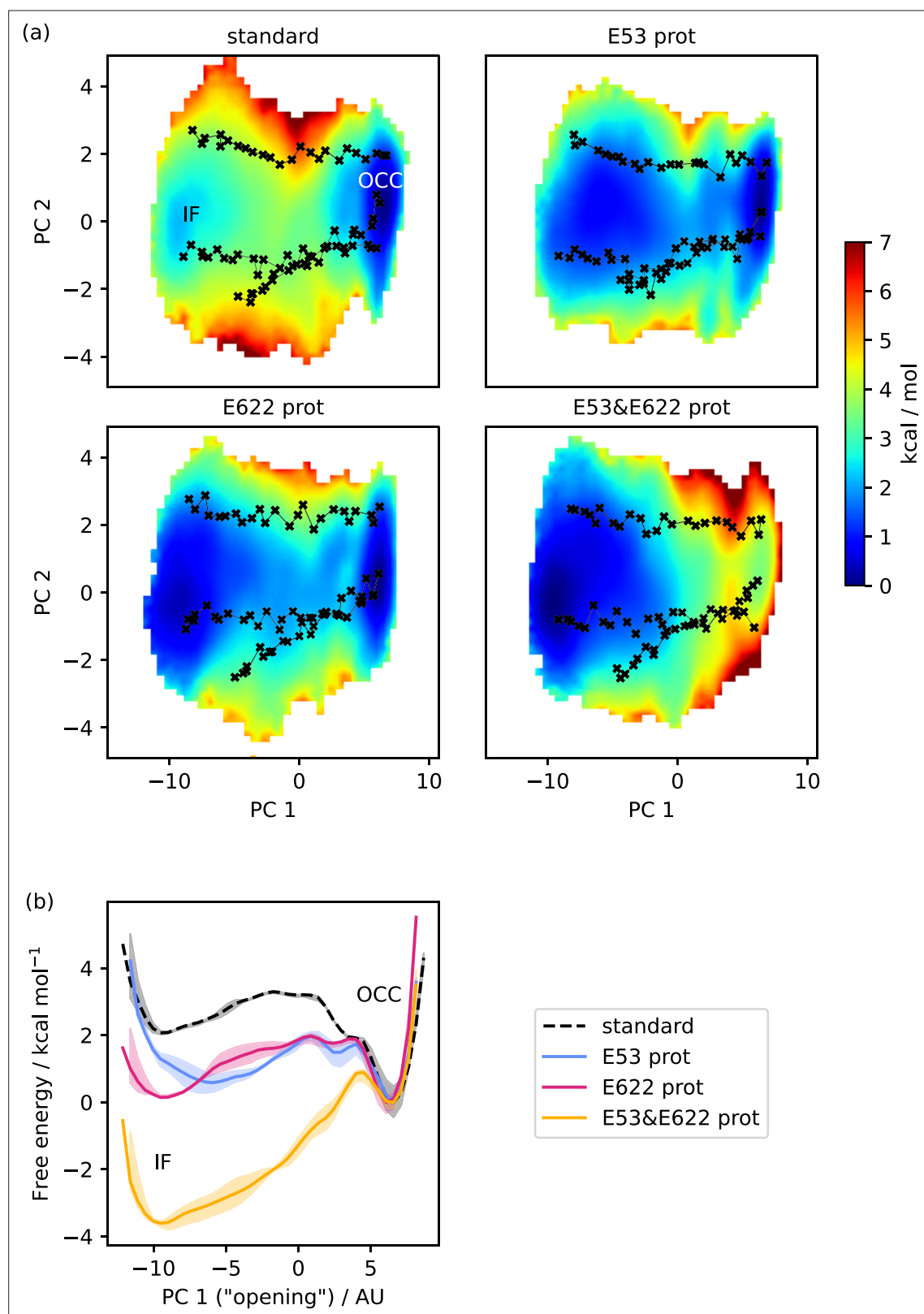


**Figure 3—figure supplement 4.** Convergence plots of all OCC→OF 2D-PMFs, shown as projections onto PC 1 including successively (increasing saturation) more data points, starting from the first frame (green) or from the last frame in reverse (orange). The PMF using all data is shown in black.



**Figure 3—figure supplement 5.** 2D-replica-exchange umbrella sampling (REUS) histograms for the OCC↔OF standard protonation state 2D-PMF, drawn as contour lines at 30% of the maximal histogram height for each window (coloured by window, from purple at occluded (OCC) to yellow at outwards-facing (OF)). Black crosses indicate the Morphing Endstates by Modelling Ensembles with iNdependent TOPologies (MEMENTO)-derived REUS starting frames. Using the lower force constant (see Materials and methods) gives good overlap in the basin regions while the transition region is undersampled. Using the higher force constant gives good overlap in the transition region while the basins are not sampled widely enough. Combining all windows results in good overlap across the 2D-collective variable (CV) space.

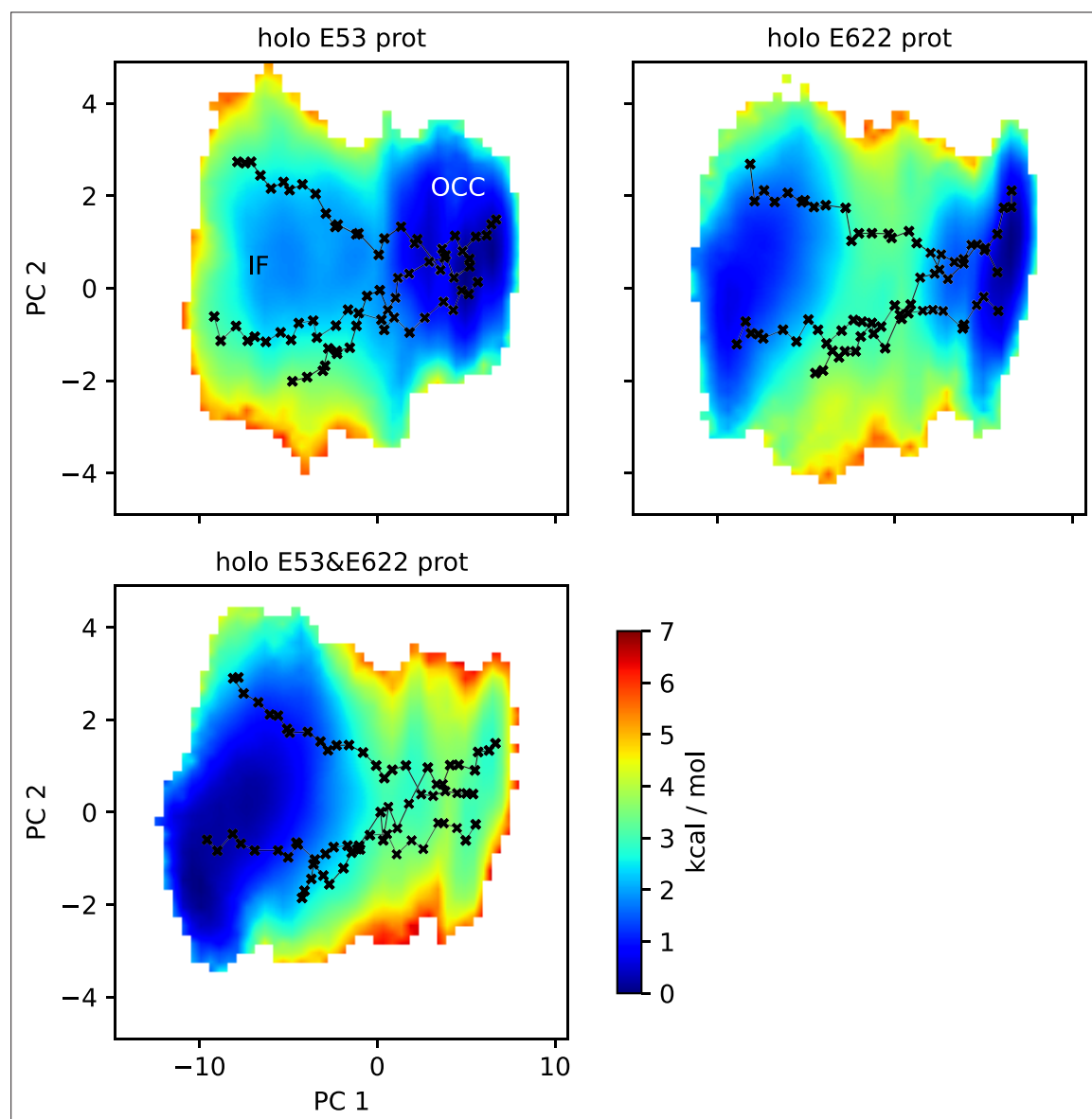




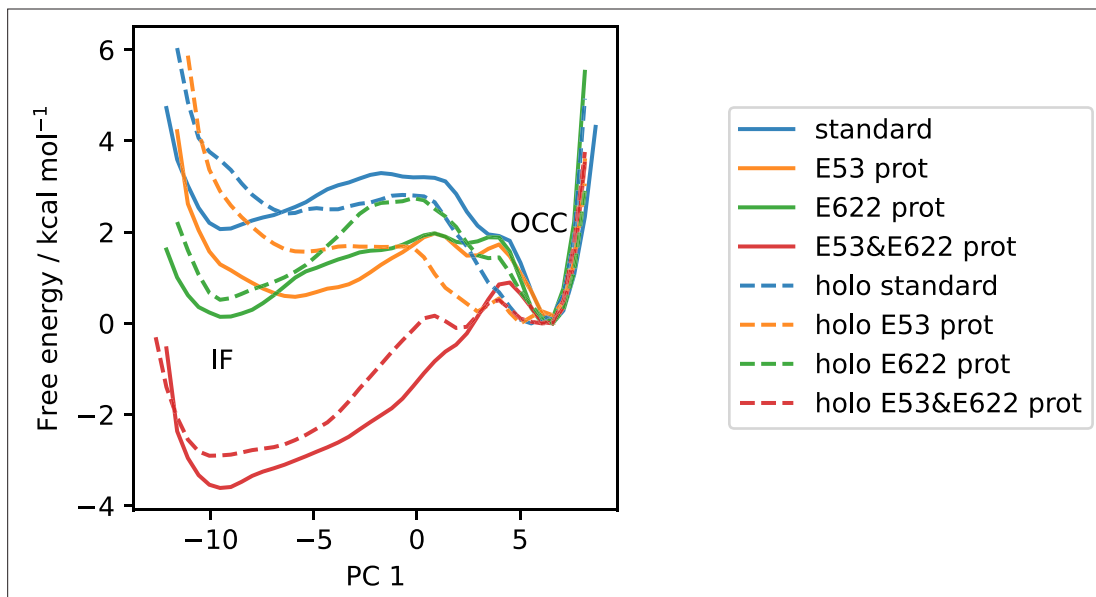
**Figure 4.** Free energy surfaces for the OCC-IF transition. (a) 2D-PMFs from replica-exchange umbrella sampling (REUS) starting with Morphing Endstates by Modelling Ensembles with iNdependent TOPologies (MEMENTO) paths, in different protonation states of candidate intracellular gate-controlling residues. (b) Projection of the 2D-PMFs in part a onto PC 1 using Boltzmann reweighting. Shaded areas indicate convergence errors as the range of PMF values for a given collective variable (CV) value obtained with the first 40%, the last 40%, and 100% of sampling included (after alignment). Figure 4 continued on next page

Figure 4 continued

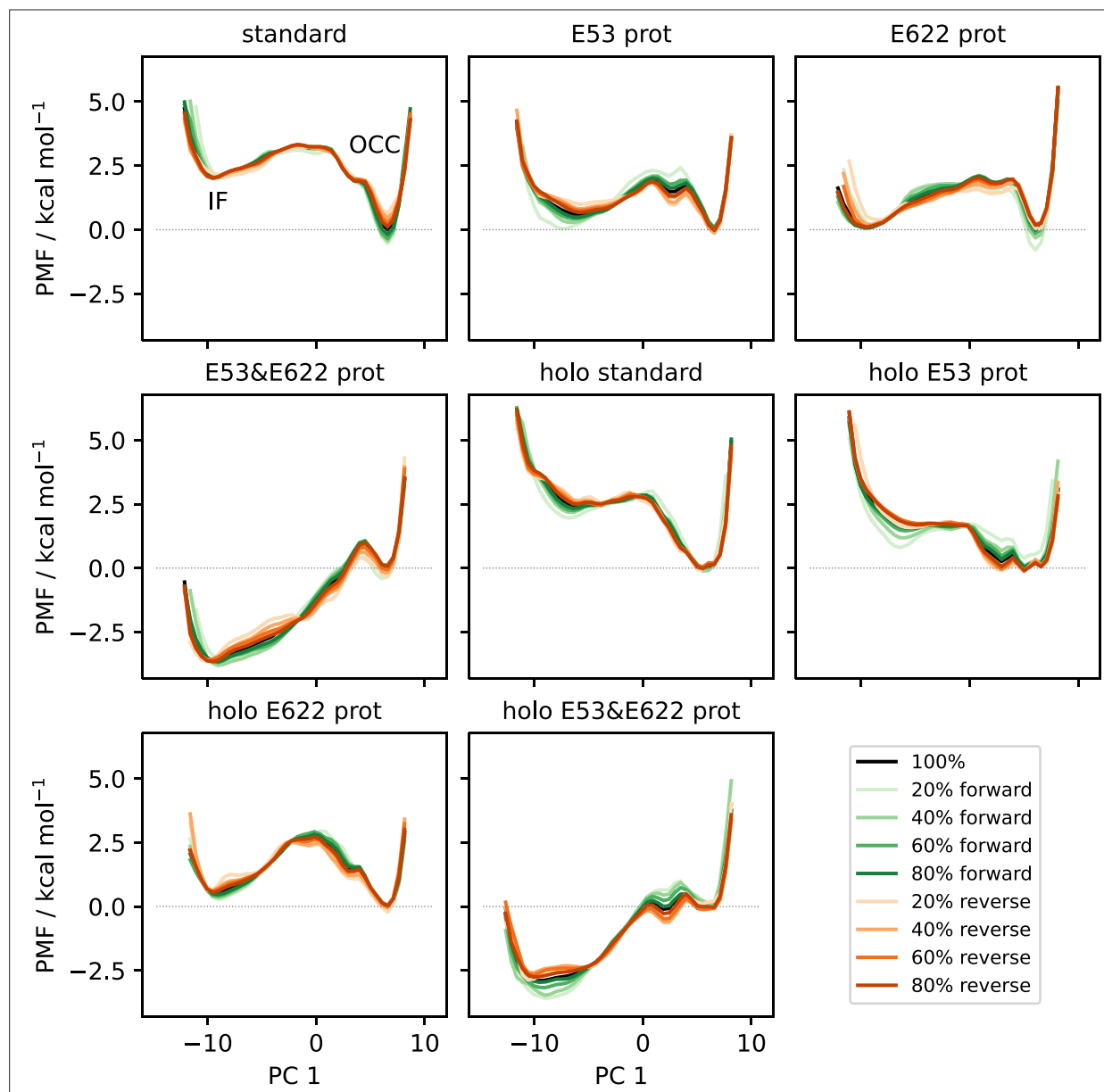
to the 100% curve). E53 and E622 protonation have additive and approximately equal effects on driving the OCC→IF transition. Note that the individual PMFs are only determined by our REUS approach up to additive constants, and are shown aligned here at the IF state for convenience of comparison.



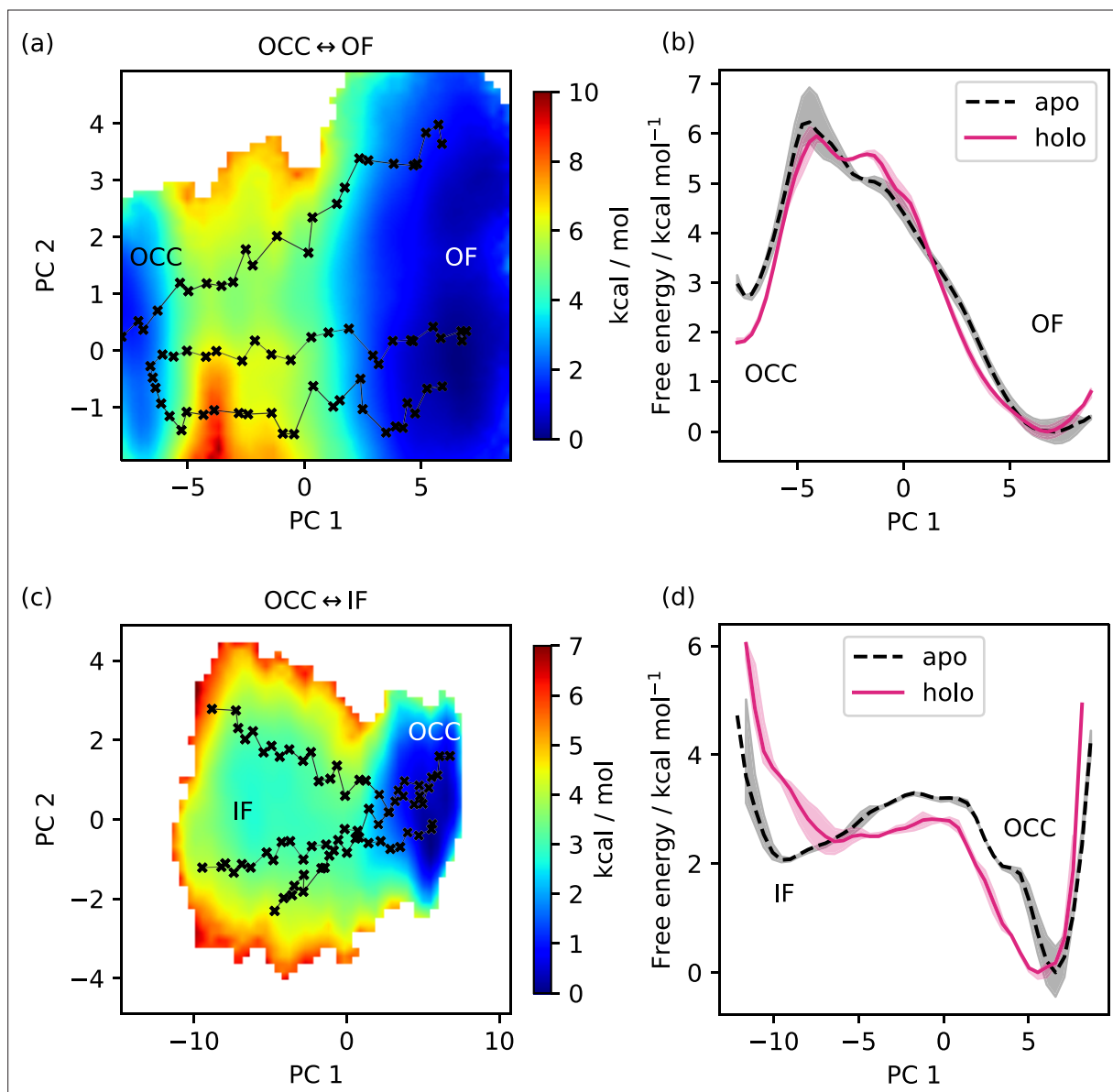
**Figure 4—figure supplement 1.** 2D-PMFs of the OCC→IF transition from replica-exchange umbrella sampling (REUS) with Morphing Endstates by Modelling Ensembles with iNdependent TOPologies (MEMENTO) paths in additional substrate-bound protonation state conditions.



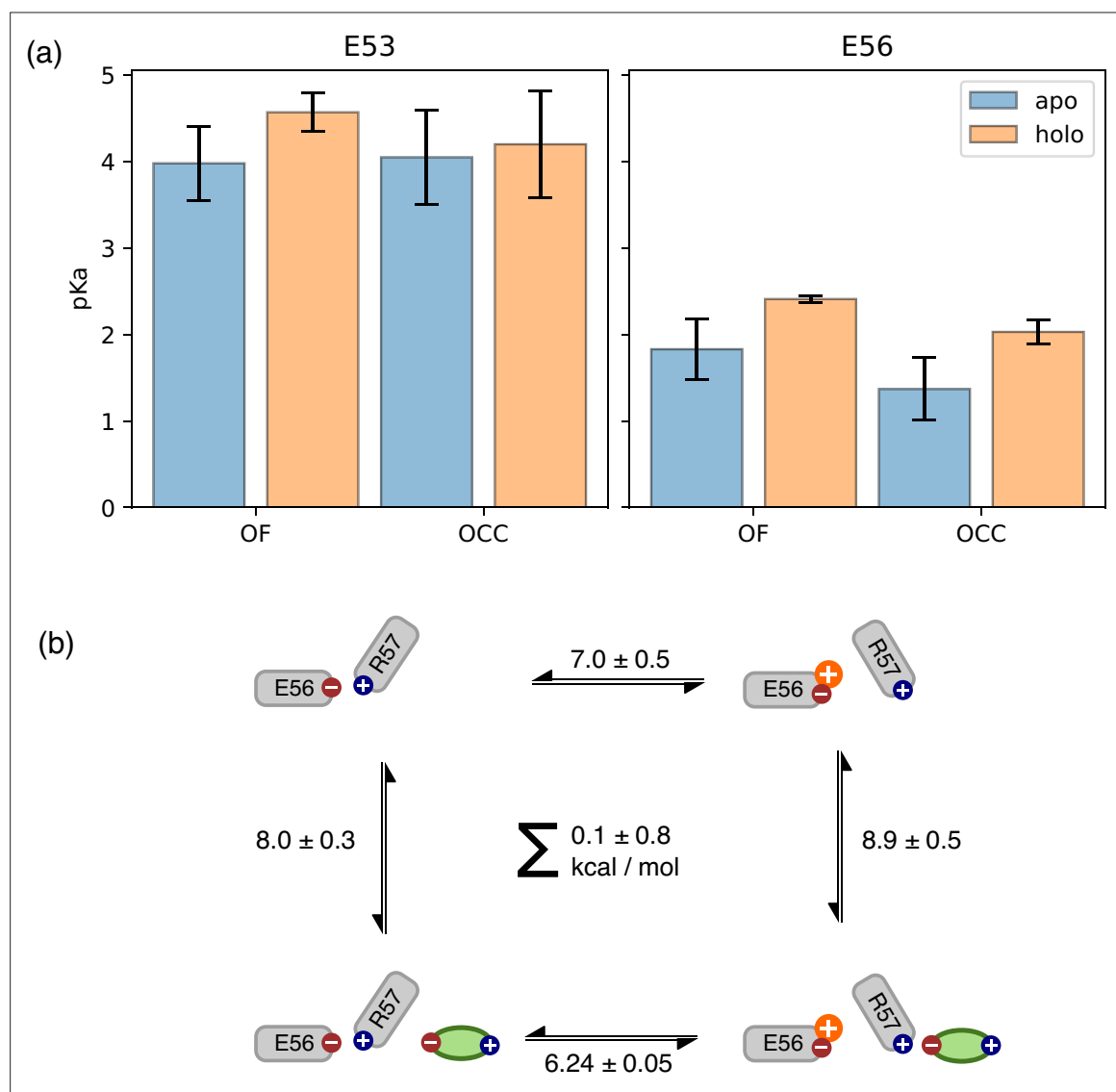
**Figure 4—figure supplement 2.** All available OCC→IF 2D-PMFs, projected onto the first collective variable (CV) (PC 1). Solid lines are apo PMFs, dashed lines are ala-phe substrate-bound and color-matched to the respective apo PMF. Note that the individual PMFs are only determined by our replica-exchange umbrella sampling (REUS) approach up to additive constants, and are shown aligned here at the inwards-facing (IF) state for convenience of comparison.



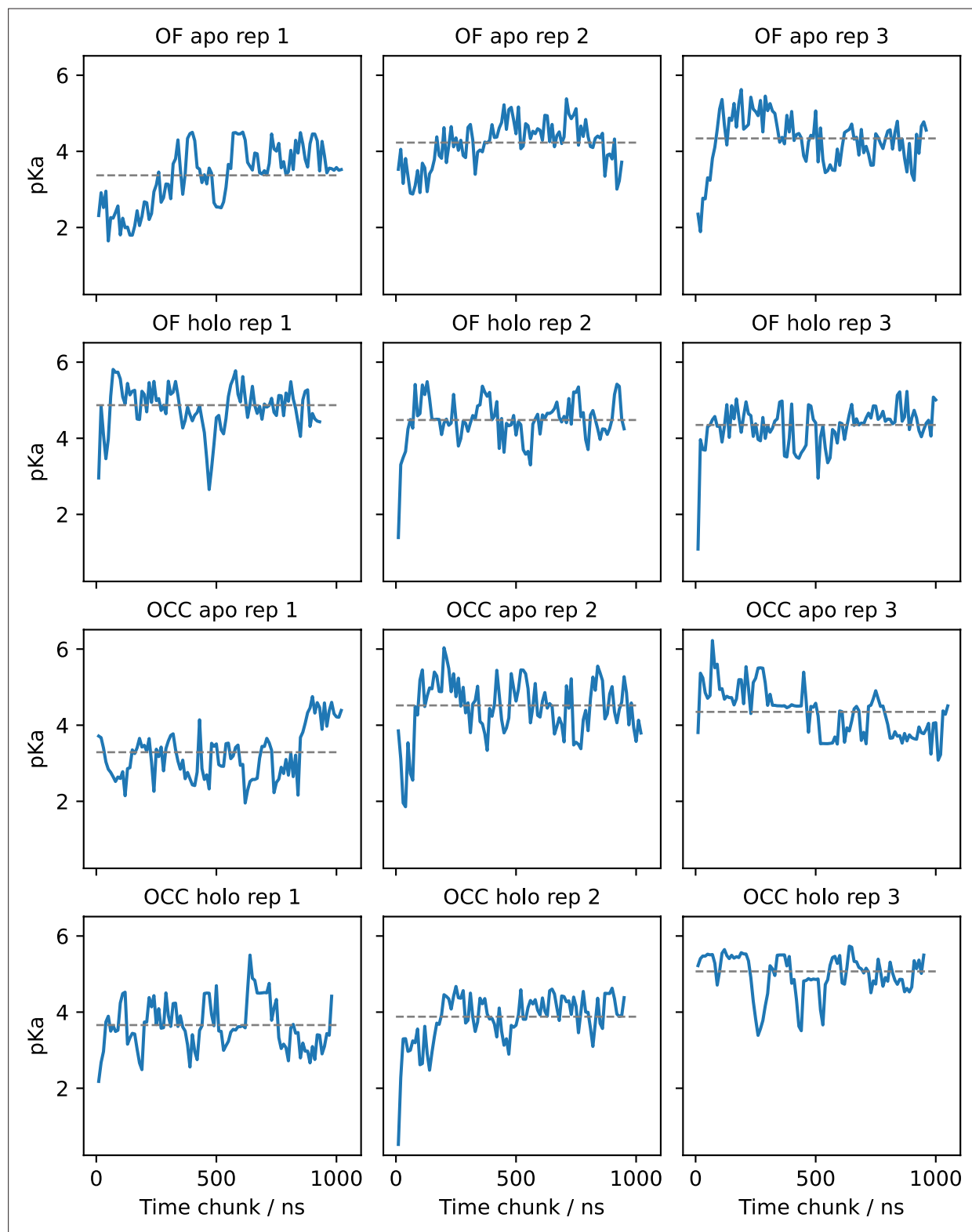
**Figure 4—figure supplement 3.** Convergence plots of all OCC→IF 2D-PMFs, shown as projections onto PC 1 including successively (increasing saturation) more data points starting from the first frame (green) or from the last frame in reverse (orange). The PMF using all data is shown in black.



**Figure 5.** Dependence of free energy surface on presence or absence of substrate. **(a)** 2D-PMF for the OCC↔OF transition from replica-exchange umbrella sampling (REUS) starting with Ala-Phe-bound PepT2 Morphing Endstates by Modelling Ensembles with iNdependent TOpologies (MEMENTO) paths. The occluded (OCC) state has an increased basin width in PC 2 (compared to **Figure 3a**), and a transition path shifted in PC 2. **(b)** Projection of the PMF from panel a onto PC 1, showing how in holo PepT2, the OCC state is stabilised by  $\approx 1$  kcal mol<sup>-1</sup>. Shaded areas indicate convergence errors as the range of PMF values for a given collective variable (CV) value obtained with the first 40 %, the last 40%, and 100% of sampling included (after alignment to the 100% curve). Note that the individual PMFs are only determined by our REUS approach up to additive constants, and are shown aligned here at the OF state for convenience of comparison. **(c)** 2D-PMF for the OCC↔IF transition from REUS starting with Ala-Phe-bound PepT2 MEMENTO paths. The structure of the inwards-facing (IF) plateau is not significantly affected, but OCC is more flexible in PC 1. **(d)** Projection of the PMF from panel c onto PC 1, showing how in holo PepT2, the OCC state has a broader basin, corresponding to intracellular-gate flexibility. Convergence error and alignment of PMFs are shown in panel b.

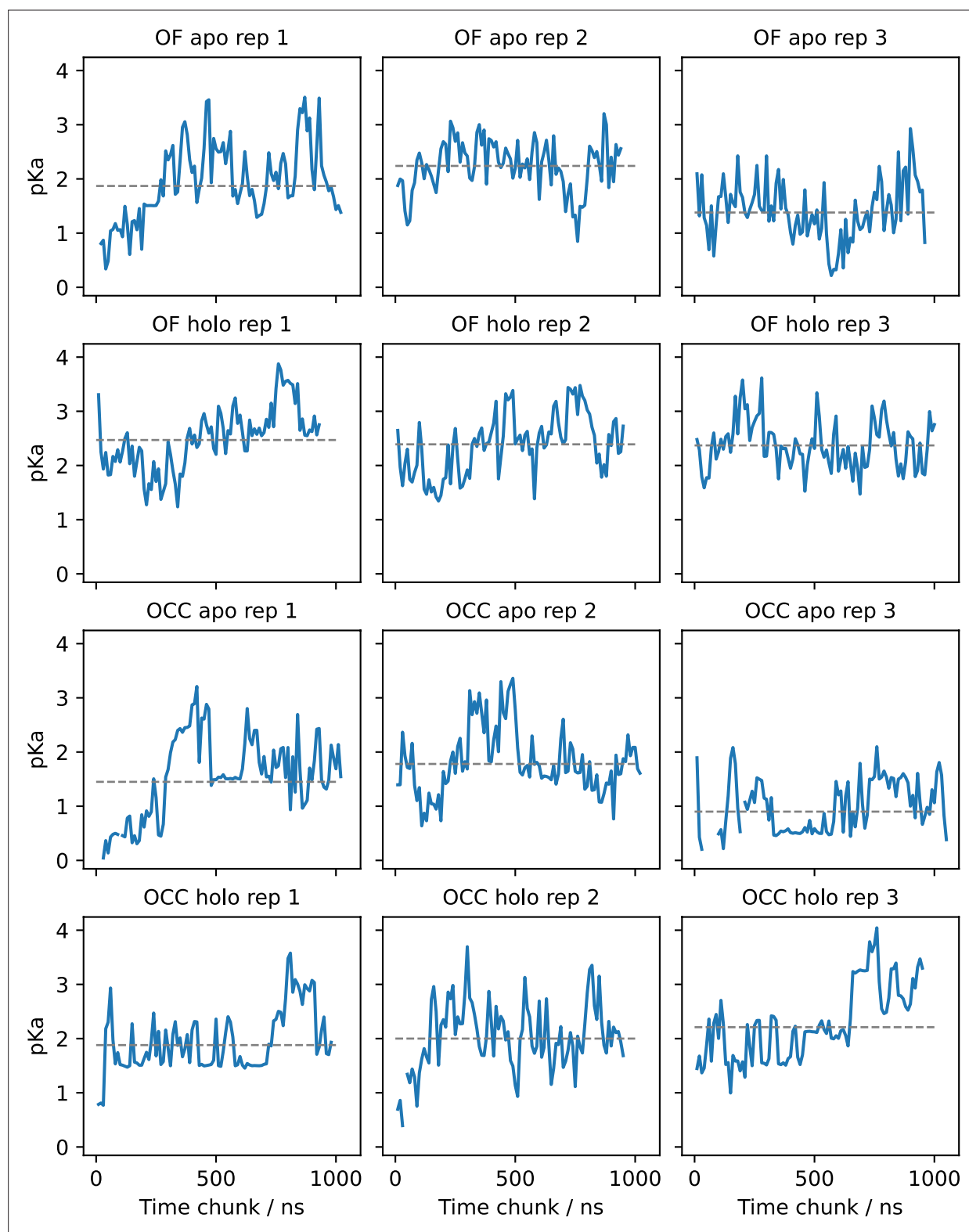


**Figure 6.** Use of constant pH simulations to investigate influence of substrate on pKa values. **(a)** E53 and E56 pKa values from constant-pH molecular dynamics (MD) simulations, in the apo and holo as well as the outwards-facing (OF) and occluded (OCC) states, estimated as mean  $\pm$  standard deviation from triplicate runs (using the full simulation data for fitting the titration curves). The presence of substrate raises the E56 pKa in either conformational state, while some effect on the E53 pKa may also exist in the OF state. **(b)** Illustration of a thermodynamic cycle of E56 protonation and Ala-Phe binding, with edges filled in via constant pH simulations (CpHMD) (converted into kcal mol<sup>-1</sup> at pH 7) for the top and bottom transitions, and absolute binding free energy (ABFE) for the left and right edges. Notably, ABFE displays a response of Ala-Phe affinity to E56 is consistent with the CpHMD results, and the cycle closes very well. The error in the cycle closure residual is estimated as a square root of the sum of squared standard deviations of the individual edges.

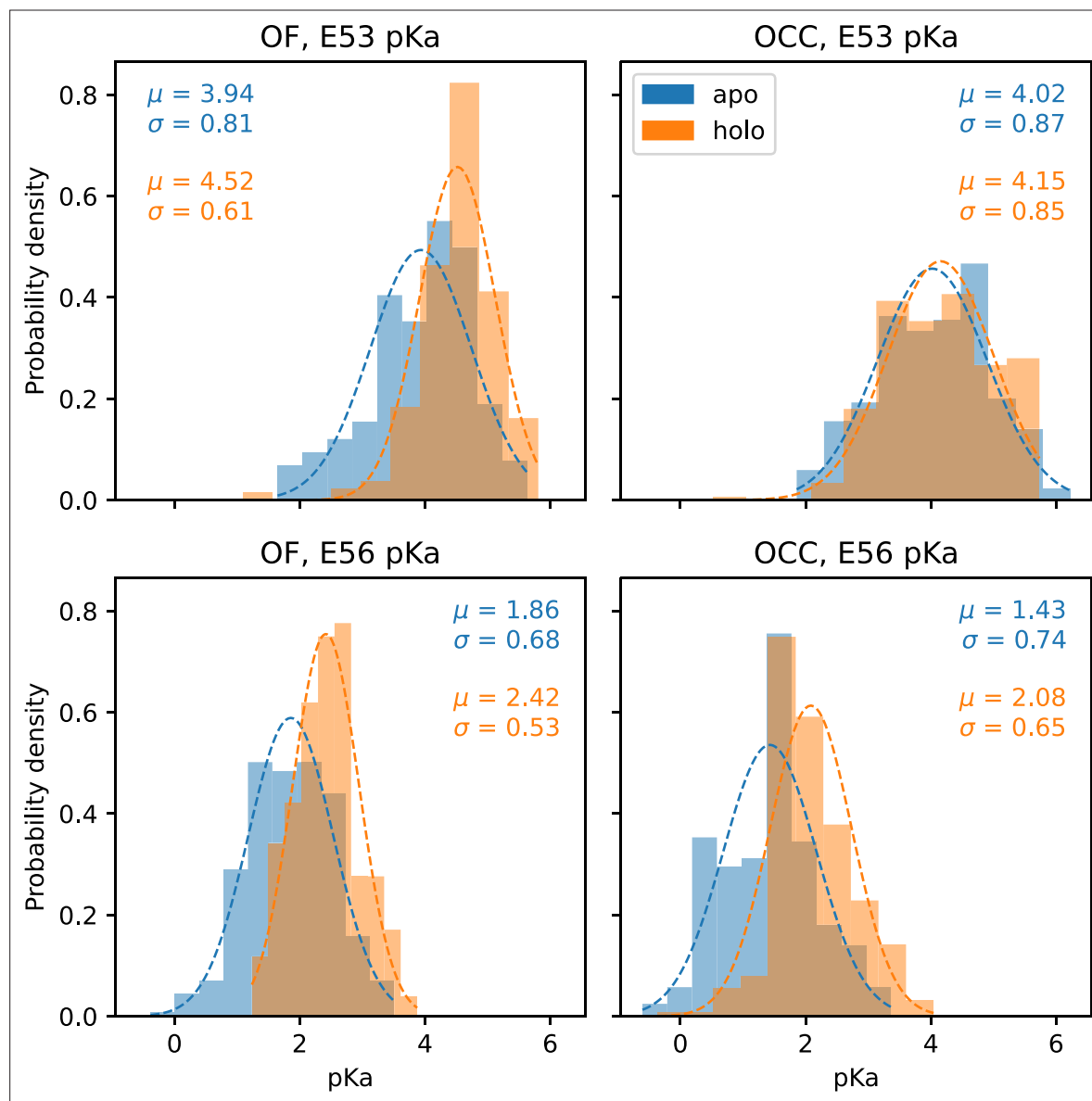


**Figure 6—figure supplement 1.** E53 pKa values estimated for (separate) successive chunks of 80 ns (10 ns per pH window) of constant pH simulations (CpHMD) via fitting the Hill equation. Grey lines indicate the pKa resulting from fitting the total simulation data. The pKa values appear highly dynamic and superimpose several relaxation timescales. Convergence seems to be achieved in most replicates but is not straightforward to assess.

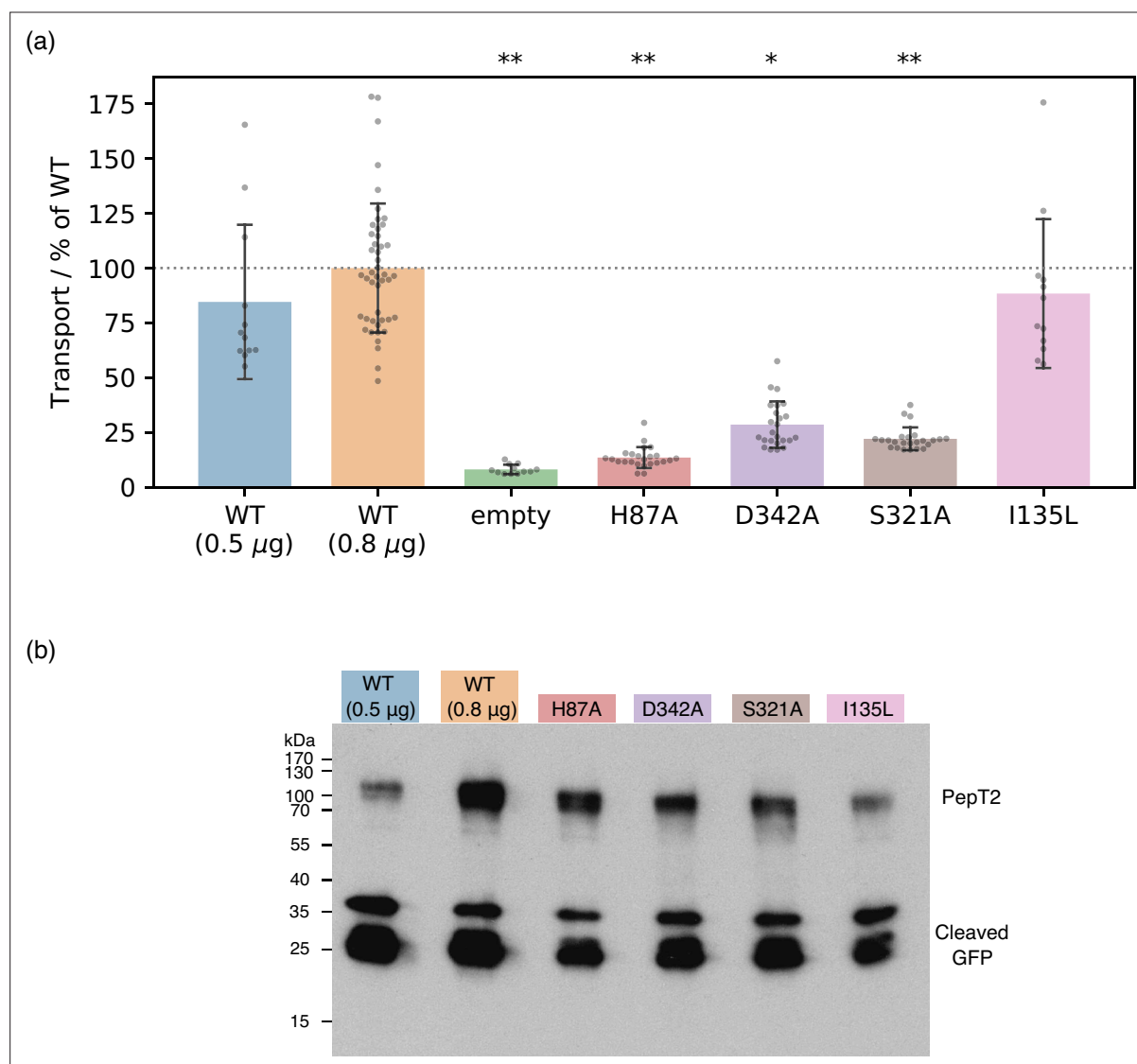




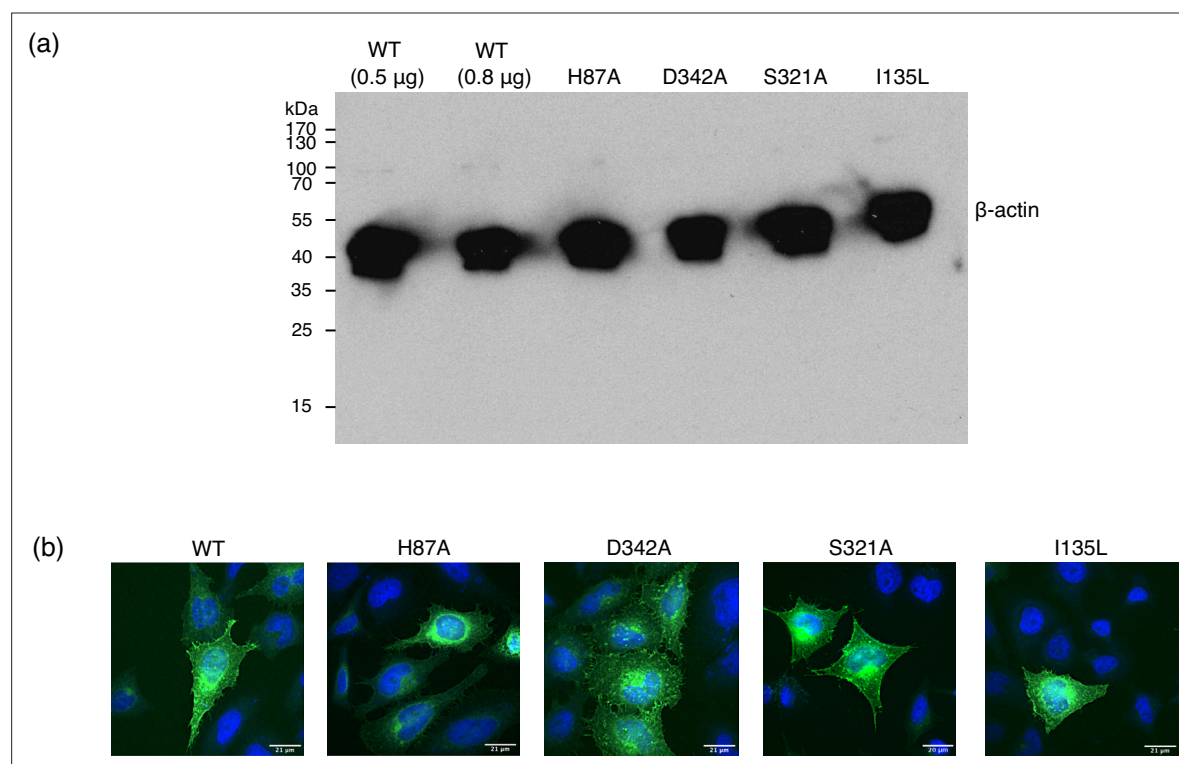
**Figure 6—figure supplement 2.** E56 pKa values estimated for (separate) successive chunks of 80 ns (10 ns per pH window) of constant pH simulations (CpHMD) via fitting the Hill equation. Grey lines indicate the pKa resulting from fitting the total simulation data. The pKa values appear highly dynamic and superimpose several relaxation timescales. Convergence seems to be achieved in most replicates but is not straightforward to assess.



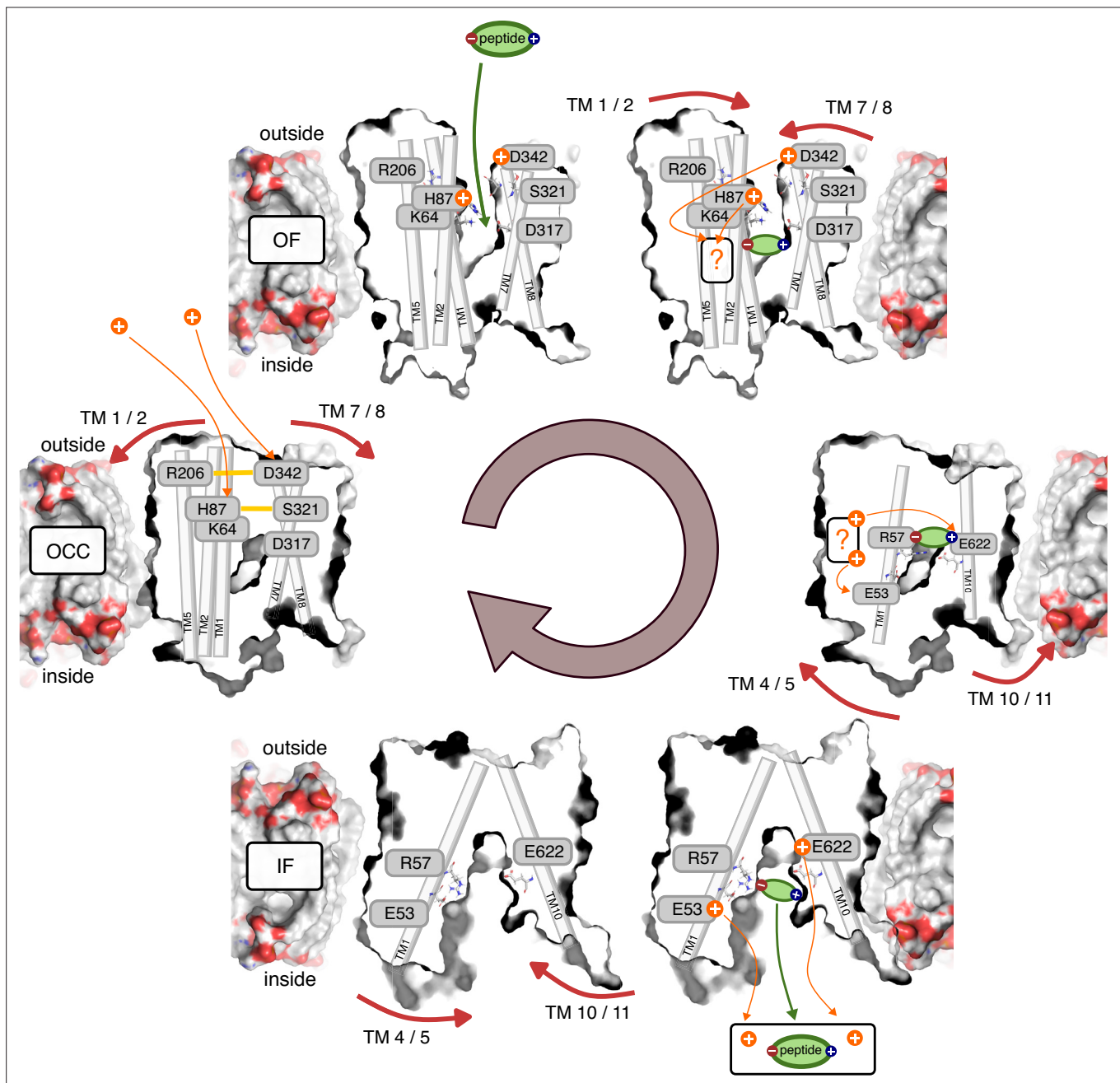
**Figure 6—figure supplement 3.** Histograms of the pKa values estimated from chunks of 80 ns (10 ns per pH window) of constant pH simulations (CpHMD), pooled for all trajectories of a given condition. Apo conditions are shown in blue, holo (ala-phe bound) are shown in orange. Gaussian fits are indicated by dashed lines and the resulting fit parameters are given in the respective color for each histogram.



**Figure 7.** Experimental validation of computational predictions. **(a)** Cell-based transport assays for PepT2 wild-type (WT) (transfected with 0.5  $\mu$ g,  $n=12$ , and 0.8  $\mu$ g,  $n=46$ , of DNA per well), empty plasmid vector ( $n=12$ ) and PepT2 H87A, D342A, S321A ( $n=24$  each) and I135L ( $n=12$ ) mutants, all transfected with 0.8  $\mu$ g of DNA. Diagram shows transport as fluorescence in post-assay lysate divided by total protein concentration, normalised to the WT (0.8  $\mu$ g) mean. Bars are mean values plus minus standard deviation, and swarm plots samples corresponding to individual wells. Single asterisk indicates  $p < 10^{-3}$ , double asterisks  $p < 10^{-4}$  significance levels for difference compared to (weaker transporting, 0.5  $\mu$ g-transfected) WT, as evaluated using a two-tailed t-test. **(b)** Western-blot showing expression levels of WT and mutant GFP-labelled PepT2, with an anti-GFP primary antibody. All mutants express at levels between the WT transfected with 0.5  $\mu$ g and 0.8  $\mu$ g plasmid DNA. Cleaved GFP is also visible at low molecular weight, at levels comparable for WT and mutants.



**Figure 7—figure supplement 1.** Control data for cell-based transport assays. (a) Loading control of the Western blot shown in **Figure 7b**, using an antibody against  $\beta$ -actin, showing even loading of the gel. (b) Fluorescence microscopy images, overlaying GFP-labelled PepT2 (green) with DAPI-labelled DNA (blue). Membrane expression is qualitatively shown for wild-type (WT) and all mutants by the thin cell outline of GFP.



**Figure 8.** Schematic overview of the PepT2 alternating-access transport cycle proposed in this work. Protons located at a question mark indicate a proton-transfer step with an as-of-yet unknown mechanism regarding intermediate residues.